

7. ТЕХНОЛОГИИ ЦИФРОВОЙ ЭКОНОМИКИ

*А. Ю. Алексей, Е.А. Хлевная, А.Е. Кузнецова,
Т.И. Карповская, Ю.Е. Карякин*

Тюменский государственный университет, г. Тюмень

УДК 004.9

АНАЛИЗ ПРОГРАММНЫХ СЕРВИСОВ ПО ОБРАБОТКЕ ПЕРВИЧНОЙ ДОКУМЕНТАЦИИ

Аннотация. Настоящая статья посвящена сравнительному анализу имеющихся программных средств обработки первичной печатной документации. В работе изучены возможности наиболее известных зарубежных и российских платформ, предоставляющих функционал для обработки первичных документов. Предложены возможные пути решения возникающих проблем ручной обработки документации.

Ключевые слова: искусственный интеллект, машинное обучение, оптическое распознавание символов, сквозная обработка, автоматизация, RPA (роботизированная автоматизация процессов), первичная документация.

Введение

Любое малое и среднее предприятие имеет бухгалтерию, и наиболее эффективным путем ведения бухгалтерского учета является автоматизация работы бухгалтерии.

Преимуществами, которые дает автоматизированная бухгалтерия, являются:

- быстрое и безошибочное заполнение реквизитов;
- оперативная обработка информационного массива;
- оперативность учета;
- представление аналитики в наглядном виде (диаграммы, графики, таблицы);

- формирование разнообразных регистров (расчетно-платежных ведомостей, кассовой книги, книг продаж и покупок);
- уменьшение количества бумаг;
- мгновенный обмен информацией между подразделениями, между руководителем и подчиненными;
- исключение арифметических ошибок;
- общение через Интернет с контролирующими государственными органами, банками;
- прозрачность и подконтрольность финансово-хозяйственной деятельности;
- быстрая реакция на изменения в законодательстве [4].

Несмотря на все преимущества использования автоматизации бухгалтерского учета, существуют следующие проблемы ее использования непосредственно в рамках предприятия:

- определение задач автоматизации, заключающиеся в необходимости определять основы функционирования предприятия и ключевые бизнес-процессы, что позволяет не выйти за рамки реальных потребностей предприятия;
- частая смена нормативных актов, устанавливающих правила учета, отчетности и налогообложения не обладают гибкостью и не способны отражать изменения внешних условий;
- недостаток квалифицированных кадров;
- перенос имеющихся данных на предприятии в новую систему [5].

Все эти проблемы могут стать причиной увеличения затрат на внедрение автоматизированной системы ведения бухгалтерского учета и отсутствия экономического эффекта от ее внедрения.

Постановка задачи

Как было отмечено, одной из значимых проблем ведения любого бизнеса являются затраты (временные и финансовые) на обработку первичной документации.

Для решения данной проблемы были осуществлены поиски подходящих сервисов по обработке первичной документации. Изучению подверглись наиболее известные российская и зарубежная платформы. Тестирования функционала сервисов проводились только на трех видах первичной документации – счет-фактурах, кредитовых авизо и банковских выписках.

Платформа «Automation Anywhere»

Данная платформа предоставляет инструмент «IQ Bot». «IQ Bot» объединяет возможности RPA с технологиями искусственного интеллекта, такими как компьютерное зрение, обработка естественного языка, нечеткая логика и машинное обучение, для автоматической классификации, извлечения и проверки информации из документов и электронных писем.

Работа «IQ Bot» делится на четыре этапа:

1. *Первичная обработка.* Улучшение качества документа с использованием методов шумоподавления, бинаризации и устранения перекоса;
2. *Классификация.* Классификация и разделение многостраничных документов для определения релевантных страниц;
3. *Экстракция.* Использование технологий машинного обучения для извлечения конкретных данных из документов;
4. *Постобработка.* Валидация с использованием правил и техник на основе искусственного интеллекта для улучшения результатов экстракции.

Данный инструмент использует оптическое распознавание символов в качестве одной из базовых технологий, используемых для извлечения информации из документа. Также «IQ Bot» использует метод сквозной обработки, чтобы измерить, сколько документов может быть обработано без вмешательства человека [1].

Сервис «Entera»

«Entera» – сервис для автоматической загрузки первичной документации в 1С, который распознает сканированные документы, Excel-файлы или фотографии [2].

Функционал «Entera» позволяет с высокой степенью точности распознавать нестандартные и нечеткие документы, хорошо определять их тип. В сервис встроены: математическая проверка таблиц, сверка с базой ЕГРЮЛ и проверка банковских реквизитов, что дополнительно минимизирует ошибки.

Большим преимуществом «Entera» с точки зрения экономии времени является возможность групповой обработки документов. У сервиса есть функция распознавания спутанных страниц и их автоматическое сопоставление в рамках одного пакета документов [3].

Практическая реализация на примере работы с сервисом «Automation Anywhere»

В качестве входных данных, подвергнутых обработке, выступили американские образцы счет-фактур, банковских выписок и кредитовых авизо, а также идентичная русская документация, самостоятельно составленная в соответствии с межгосударственным стандартом (ГОСТом).

Далее представлены образцы российской (рис. 1) и американской (рис. 2) счет-фактур для наглядной демонстрации структуры документов.

СЧЕТ-ФАКТУРА № 55 от "26" апреля 2020 г.
 ИСПРАВЛЕНИЕ № -- от --
 Продавец: Акционерное общество «Номер 1»
 Адрес: адрес1
 ИНН/КПП продавца: 6327743889/234723489
 Грузоотправитель и его адрес: -
 Грузополучатель и его адрес: -
 К платежно-расчетному документу: № 240 от 20.12.2019, №245 от 24.11.2019
 Покупатель: ООО ИП ИНВЕСТ
 Адрес: 625048, г. Тюмень, ул. Центральная, д.25, кв.16
 ИНН/КПП покупателя: 2367848778/727872384
 Валюта: наименование, код Российской рубль, 643
 Идентификатор государственного контракта, договора (соглашения) (при наличии):

Наименование товара (описание выполненных работ, оказанных услуг), имущественного права	Код вида товара	Единица измерения		Количество (объем)	Цена (тариф) за единицу измерения	Стоимость товаров (работ, услуг), имущественных прав без налога - всего	В том числе сумма акциза	Налоговая ставка	Сумма налога, подлежащая уплате покупателем	Стоимость товаров (работ, услуг), имущественных прав с налогом - всего	Страна происхождения товара		Регистрационный номер таможенной декларации
		код	условное обозначение (национальное)								цифровой код	краткое наименование	
1	1a	2	2a	3	4	5	6	7	8	9	10	10a	11
Плитка1	.	796	шт/pcs	3	175.23	525.69	без акциза	20	105.138	630.828	-	-	-
Плитка2	.	796	шт/pcs	2	230.22	460.44	без акциза	20	92.088	552.528	-	-	-
Плитка3	.	796	шт/pcs	1	340	340	без акциза	20	68	408	-	-	-
Всего к оплате						1326.13		X	265.226	1591.356			

Руководитель организации или иное уполномоченное лицо _____ (подпись) (ф.и.о.)
 Главный бухгалтер или иное уполномоченное лицо _____ (подпись) (ф.и.о.)

Индивидуальный предприниматель или иное уполномоченное лицо _____ (подпись) (ф.и.о.)
 (реквизиты свидетельства о государственной регистрации индивидуального предпринимателя)

Рис. 1. Образец российской счет-фактуры

INVOICE

Invoice no.	10265	Purchase Order	26522
Invoice Date	22 Aug 2018	Terms	2% 10 Net 30

Sold to Ultron Inc. 1801 Toldeo Beach Rd #1189 La Salle, MI 48145	Ship to Ultron Inc. - Receiving Dock 6395 N Russel Rd Oak Harbor, OH 43449
---	--

1019

Quantity	Item no.	Description	Unit Price	Tax	Amount
1	1001	Widget Type A	22.00	G	22.00
1	1007	Bracket - Widget Type C	36.75	G	36.75
3	1010	Sprocket - Med	18.95	G	56.85
10	1009	Sprocket - Small	8.90	G	89.00

Subtotal	204.60
Tax1 (GST) @ 7%	14.32

Invoice Amount \$ 218.92

Рис. 2. Образец американской счет-фактуры

Работа по созданию бота велась в разделе «COGNITIVE AUTOMATION. IQ BOT». По нажатию на кнопку «Создать новый экземпляр» открылся «пункт управления», в котором есть несколько разделов.

В разделе «Основная информация» необходимыми полями для заполнения являются: *имя* создаваемого экземпляра, *тип* документа и *основной язык* документов. Выпадающий список «*тип документа*» предоставляет пользователю возможность выбрать один из пяти типов документов: банковские выписки, кредитовое авизо, счета-фактуры, заказы, счета за коммунальные услуги. Поля для извлечения этих типов документов генерируются автоматически. Такая возможность предоставляется носителям таких языков, как английский, немецкий, французский, испанский и итальянский. Российским пользователям для анализа необходимой документации требуется в указанном выпадающем списке выбрать пункт «Другой». Далее в поле «Основной язык

документа» выбрать русский язык и вписать наименование документа в поле «Пользовательский тип документа».

Следующий раздел посвящен импорту документов. Максимально допустимое число загружаемых документов – 150.

Затем идет раздел «Поля для извлечения». Если пользователем был выбран один из пяти предоставленных типов документов, то поля для извлечения сформируются автоматически. Если же в выпадающем списке «Тип документа» был выбран пункт «Другой», то в этом разделе пользователь будет вынужден самостоятельно вводить названия полей для извлечения данных из документа.

Последний раздел, являющийся необязательным, это раздел под названием «Таблица/повторяющиеся поля раздела». Тут можно указать таблицу, то есть поля, которые представлены не в единственном экземпляре и не являющиеся уникальными.

Следующим шагом является нажатие на кнопку «Создать экземпляр и проанализировать», после чего начинается анализ документов. Анализ 10 документов на практике занимает 1-2 минуты. В процессе анализа учебный экземпляр завершает классификацию документов, которые в дальнейшем будут предоставлены пользователю в качестве полученного результата. Документы на конечном этапе анализа делятся по группам. Группы формируются автоматически, исходя из внешних и содержательных различий. Результатом анализа будут являться созданные экземпляры «IQ Bot» для каждого типа групп.

На следующем этапе пользователю предоставляется результат анализа. Пользователь может просмотреть документ каждой сформировавшейся группы, самостоятельно оценить корректность извлеченных данных и в случае какой-либо ошибки отредактировать полученные значения следующим образом: выбрать каждое поле, которое указано на левой панели, чтобы убедиться, что метка и значение, которое представлено на центральной панели, совпадают с документом, который находится на правой панели; провести предварительный просмотр сопоставления полей и меток после проверки всех полей; сохранить

выбора вручную полей документа для распознавания, в отличие от сервиса Automation Anywhere.

Далее было решено продолжить исследование этого российского сервиса, загрузив российские счет-фактуры. После загрузки счет-фактур, сервис распознал этот тип документа (рис. 4).

<input type="checkbox"/>	Статус	Плательщик	Поставщик	Тип документа	Номер документа	Дата документа	Дата загрузки	↑	⚙
<input type="checkbox"/>	Распознан	ООО "МПИ"	ООО "ЛЕРУА МЕРЛЕН ВОСТОК"	УПД	031/2020-0003110	02.02.2020	02.05.2020		
<input type="checkbox"/>	Распознан	ООО "МПИ"	ООО "ЛЕРУА МЕРЛЕН ВОСТОК"	УПД	031/2020-0004019	09.02.2020	02.05.2020		
<input type="checkbox"/>	Распознан	ООО "МПИ"	АКЦИОНЕРНОЕ ОБЩЕСТВО "ЭР-ТЕЛЕКОМ ХОЛДИНГ"	Счёт-фактура	2505/18	31.01.2020	02.05.2020		
<input type="checkbox"/>	Распознан	ООО "МПИ"	ООО "ГАЗПРОМНЕФТЬ - КОРПОРАТИВНЫЕ ПРОДАЖИ"	УПД	CSC0000000072233	31.01.2020	02.05.2020		
<input type="checkbox"/>	Распознан	ООО "МПИ"	ООО "ГАЗПРОМНЕФТЬ - КОРПОРАТИВНЫЕ ПРОДАЖИ"	УПД	CSC0000000072273	31.01.2020	02.05.2020		
<input type="checkbox"/>	Распознан	ООО "МПИ"	ООО "ЛЕРУА МЕРЛЕН ВОСТОК"	УПД	031/2020-0004492	12.02.2020	02.05.2020		
<input type="checkbox"/>	Распознан	ООО "МПИ"	ООО ППО "ЗАПСИБГАЗЭНЕРГОКОМПЛЕКТ"	Счёт-фактура	10/01	28.01.2020	02.05.2020		
<input type="checkbox"/>	Распознан	ООО "МПИ"	ООО "ГАЗПРОМНЕФТЬ - КОРПОРАТИВНЫЕ ПРОДАЖИ"	Счёт-фактура	CSC0000001017135	30.11.2019	02.05.2020		
<input type="checkbox"/>	Распознан	ООО "МПИ"	ООО ППО "ЗАПСИБГАЗЭНЕРГОКОМПЛЕКТ"	Счёт-фактура	10/11	25.11.2019	02.05.2020		
<input type="checkbox"/>	Распознан	ООО "МПИ"	ООО "ГАЗПРОМНЕФТЬ - КОРПОРАТИВНЫЕ ПРОДАЖИ"	УПД	CSC0000000003108	31.01.2020	02.05.2020		

Рис. 4. Распознанные документы

Результаты обработки первичной документации в сервисах «Automation Anywhere» и «Энтера»

1. Результаты обработки первичной российской документации в сервисе «Automation Anywhere»:

При создании ботов для обработки первичной документации на основе счет фактур, банковских выписок и кредитовых авизо во всех трех случаях была сформирована одна группа документов, так как обрабатываемые образцы были предоставлены в едином виде. Точность извлечения варьируется от 20% до 43%, так как при извлечении данных из определенного количества полей корректно обрабатывалась лишь некоторая часть данных. В процессе корректировки полученных данных путем обучения экземпляра точность извлечения может быть достигнута 100%, что и было получено практическим путем.

Преимуществом в данном случае является то, что образцы обрабатываемых документов имеют одну структуру, т. к. сформированы они в соответствии с ГОСТом, что позволяет с легкостью корректно извлекать данные из всех полей.

Существенным недостатком при создании экземпляра «IQ Bot» является тот факт, что необходимо прописывать наименования полей вручную, так как эта платформа не предоставляет список возможных полей документа, кроме как для американской документации.

2. Результаты обработки первичной американской документации в сервисе «Automation Anywhere»:

Во всех трех случаях (при обработке счет-фактур, банковских выписок и кредитовых авизо) были сформированы две группы документов. Точность извлечения из той и другой группы варьируется от 88% до 100%. В процессе обучения экземпляра точность достигает значения 100%.

Достоинством является то, что при создании экземпляра выбор полей предоставляется автоматически, что заметно экономит время при создании экземпляра «IQ Bot». Также при обучении бота неповторяющиеся поля сразу же извлекаются безошибочно в любой группе документов, а значит, платформа уже хорошо обучена для этого типа документов, что является значительным плюсом.

Существенный недостаток – при обучении бота извлекать повторяющиеся поля из таблицы возникли небольшие сложности, которые впоследствии были устранены, а именно: таблицы в документах были расположены в разных местах (выше или ниже положенного места).

3. Результаты обработки первичной российской документации в сервисе «Энтера»:

При работе с этой платформой были загружены 10 документов – российские счет-фактуры. Точность извлечения данных составила 100%. С американскими документами сервис не работает.

Минусами этой платформы являются: платформа не предназначена для бесплатного использования; на данной платформе не реализовано распознавание всех типов документов.

Сравнительный анализ результатов обработки первичных документов

Опираясь на результаты обработки первой документации в сервисе «Automation Anywhere», была составлена сводная таблица по результатам проведенного исследования (табл. 1).

Таблица 1. Результаты анализа

Критерий	Российские образцы	Американские образцы
Точность извлечения	20 – 43%	88 – 100%
Время извлечения	2–3 мин	2–3 мин
Удобство использования	3/5	5/5

Сравнительный анализ по результатам обработки документов на платформе «Enter» не приводился, т. к. данный сервис ориентирован исключительно на работу с российской документацией, что является существенным минусом в сравнении с платформой «Automation Anywhere». Однако в процессе изучения сервиса «Enter» было выявлено, что платформа обладает всем необходимым функционалом, позволяющим извлекать данные из полей документов с уровнем точности, равным 100%, что является значимым преимуществом. В свою очередь, количество типов документов, из которых платформа позволяет извлекать данные, слишком мало, а также пользование сервисом может осуществляться исключительно на платной основе.

Заключение

Зарубежный сервис «Automation Anywhere» хорошо адаптирован под нужды граждан своего государства, и функционал сервиса плохо «заточен» на работу с документацией, отличной от американских документов. Единственный

способ решения проблемы обработки в данном случае – стандартизация нынешней документации государства по типу документов США.

Российский сервис предназначен для работы исключительно с ограниченным числом отечественной документацией. Существенным преимуществом является тот факт, что обработка документов осуществляется идеально, никакие самостоятельные доработки не требуются, в отличие от обработки документации в сервисе «Automation Anywhere». Относительно зарубежного сервиса функционал «Entera» очень ограничен. Эта платформа будет бесполезна в использовании за рубежом, а обработка может осуществляться только для очень ограниченного числа типов документации.

Таким образом, были сформулированы следующие возможные пути решения проблемы обработки первичной документации: расширение функционала рассмотренного отечественного сервиса; стандартизация нынешней документации государства по типу документов США.

В результате все предприятия могут существенно облегчить процесс обработки первичных документов, и тем самым сэкономить свои финансовые и временные ресурсы.

СПИСОК ЛИТЕРАТУРЫ

1. Automation Anywhere: Global RPA Solutions. – URL: <https://www.automationanywhere.com/> (дата обращения: 24.04.2021).
2. Entera. – URL: <https://cutt.ly/skxWzHa> (дата обращения: 24.04.2021).
3. Обзор российских сервисов по обработке первичной документации. – URL: <https://www.klerk.ru/buh/articles/486972/> (дата обращения: 22.04.2021).
4. Автоматизация бухгалтерского учета: как выбрать и внедрить систему. – URL: <https://lab-automat.ru/avtomatizatsiya/avtomatizatsiya-buhgalterskogo-ucheta-kak-vybrat-i-vnedrit-sistemu.html> (дата обращения: 22.04.2021).
5. Современные проблемы в бухгалтерском учете. – URL: <http://journalpro.ru/articles/sovremennye-problemy-v-bukhgalterskom-uchete/> (дата обращения: 22.04.2021).