

DOI 10.21672/2074-1707.2021.53.1.081-089
УДК 004[01+9.2]

АВТОМАТИЗАЦИЯ ЭЛЕКТРОННО-ПЕЧАТНОГО ДЕЛОПРОИЗВОДСТВА С ИДЕНТИФИКАЦИЕЙ И ВЕРИФИКАЦИЕЙ ДОКУМЕНТОВ

Статья поступила в редакцию 26.04.2021, в окончательном варианте – 16.05.2021.

Петров Дмитрий Никифорович, Санкт-Петербургский государственный технологический институт (технический университет), 190013, Российская Федерация, г. Санкт-Петербург, пр. Московский, 24-26/49 литер А,
кандидат технических наук, ORCID 0000-0003-0083-784X, e-mail: petrov.dmitry@technolog.edu.ru

Луцко Андрей Николаевич, Санкт-Петербургский государственный технологический институт (технический университет), 190013, Российская Федерация, г. Санкт-Петербург, пр. Московский, 24-26/49 литер А,
кандидат технических наук, доцент, ORCID 0000-0001-5350-1595, e-mail: lutsko@technolog.edu.ru

Рассмотрена и проанализирована проблематика электронно-печатного делопроизводства, идентификации и верификации документов. Обоснована техническая целесообразность применения алгоритмов распознавания рукописного текста, искусственного интеллекта и криптографии при электронно-печатном делопроизводстве. Предложен способ автоматизированной обработки документа на бумажном носителе и проверки на подлинность его электронного аналога. Подробно изложен алгоритм альтернативного четырехстадийного жизненного цикла документа и методика его верификации при сокращении ручных операций и машинном вводе. Представлена структура интегрированной подсистемы машинного ввода и верификации документов, используемая в рамках Единой информационной системы «Электронный Университет» образовательного учреждения. Получены результаты автоматизированной обработки и защиты документа с обучением многослойной сверточной нейронной сети на примере ведомостей промежуточной аттестации. Доказана возможность применения изложенной методологии к документам строгой отчетности фиксированной структуры с рукописным заполнением.

Ключевые слова: электронно-печатный документооборот, идентификация, верификация, распознавание образов, нейронная сеть

AUTOMATION OF ELECTRONIC-PRINTING OFFICE WORK WITH THE DOCUMENTS IDENTIFICATION AND VERIFICATION

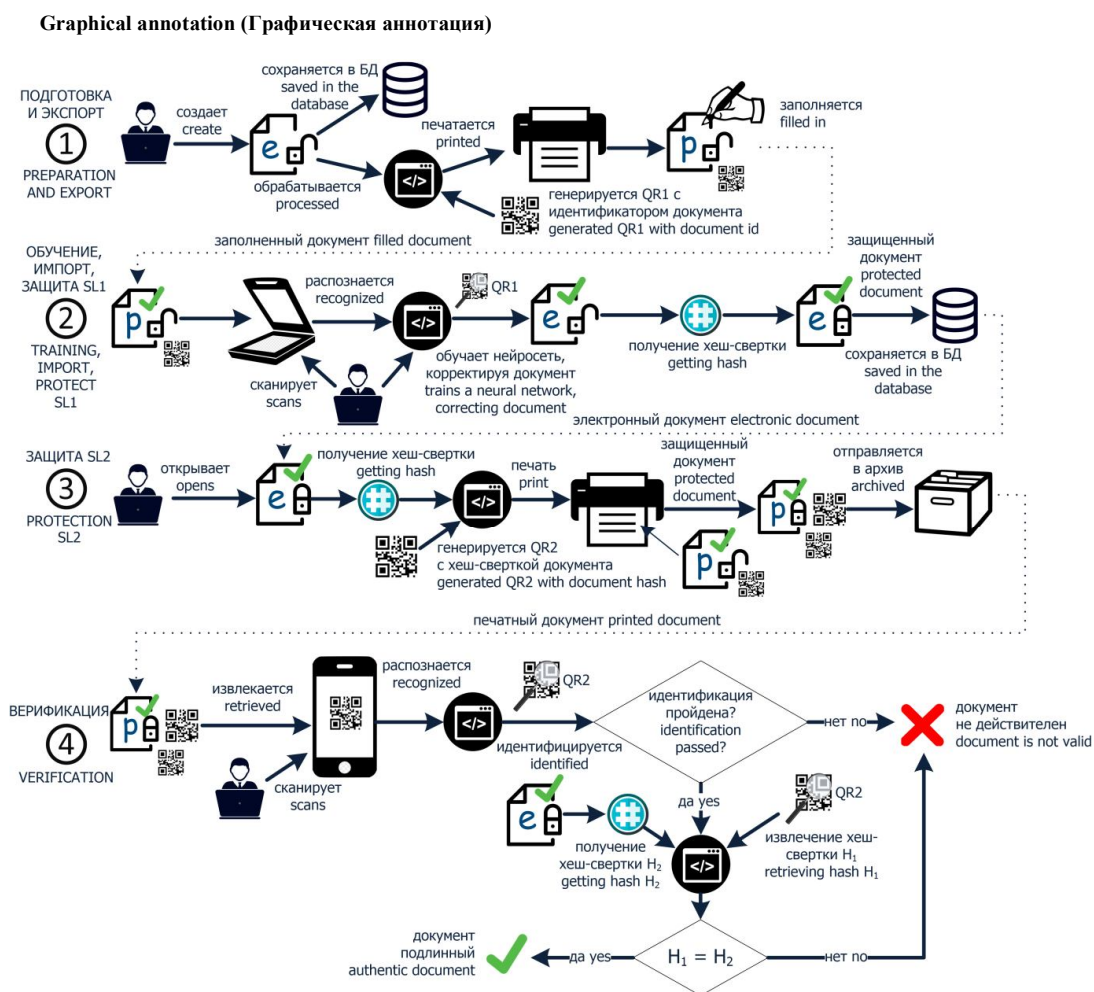
The article was received by the editorial board on 26.04.2021, in the final version – 16.05.2021.

Petrov Dmitriy N., Saint-Petersburg State Institute of Technology (technical university), 24-26/49 lit A Moskovskiy Ave., Saint Petersburg, 190013, Russian Federation,
Cand. Sci. (Engineering), ORCID 0000-0003-0083-784X, e-mail: petrov.dmitry@technolog.edu.ru

Lutsko Andrey N., Saint-Petersburg State Institute of Technology (technical university), 24-26/49 lit A Moskovskiy Ave., Saint Petersburg, 190013, Russian Federation,
Cand. Sci. (Engineering), Associate Professor, ORCID 0000-0001-5350-1595, e-mail: lutsko@technolog.edu.ru

In the article the problem of electronic-printing office work, the documents identification and verification has been reviewed and analyzed. The technical feasibility of using algorithms for handwriting recognition, artificial intelligence and cryptography in electronic-printing office work has been substantiated. Authors are suggest an automated processing method of a paper document verification of its electronic counterpart authenticity. The algorithm of the alternative four-stage life cycle of the document and the method of its verification with the manual operations reduction and machine input are presented in detail. The structure of the integrated subsystem of machine input and documents verification used in the Edinaja informacionnaja sistema «Elektronniy Universitet» of an educational institution is presented. The results of the document automated processing and protection with training a multilayer convolutional neural network using the example of intermediate certification sheets were obtained. The possibility of applying the described methodology to strict reporting of a fixed structure documents with handwritten filling has been proved.

Keywords: electronic-printing office work, identification, verification, pattern recognition, neural network



Введение. С начала XXI столетия стремительно развивается тренд электронного делопроизводства. С ростом объемов данных и ужесточением требований к скорости, доступности и качеству их обработки, традиционный бумажный документооборот признан неэффективным. Однако не стоит отрицать чрезвычайную устойчивость бумажных и подобных им носителей к самым негативным воздействиям природы или человека. Действительно, выражение «рукописи не горят» имеет под собой физическую основу. Работа с бумажными документами не требует специальных устройств чтения, а время жизни базовой лингво-знаковой системы составляет тысячи лет. Электронные данные утрачивают читаемость не только вследствие экстремального физического воздействия на их носители, но и из-за их относительно короткого срока службы. Кроме того, время жизненного цикла компьютерных баз данных (БД) ограничено перманентным развитием технологий энергонезависимых запоминающих устройств с отказом от устаревших технологий, что также препятствует прочтению архивных данных, не приведенных к новому стандарту и формату записи. Данные на электронных носителях также подвержены искажению вследствие системных сбоев, вирусных атак.

Благодаря преимуществам, достигаемым при электронном документообороте, и преимуществам при использовании бумажных носителей, электронно-печатный способ делопроизводства, принятый большинством предприятий, будет использоваться еще достаточно продолжительное время. Это время продлевают финансовые, технические и организационные сложности использования и обслуживания дорогостоящих сертифицированных информационных систем повышенной надежности и средств безопасности данных, а также высокая стоимость бланков документов с технологической, физико-химической или полиграфической защитой [1].

Наиболее часто используемой поэтапной схемой электронно-печатного документооборота является формирование бланка электронного документа, его печать с последующим рукописным заполнением, ручной ввод данных в БД с заполненного бланка, сканирование бумажной версии документа и сопряжение ее с электронной версией, архивирование.

Поэтому актуальной является задача повышения скорости и качества ввода (оцифровки) рукописных данных с наложением защиты на электронную и печатную версии документа, что приводит их к жесткому соответствию и неподдельности.

Постановка задачи и аналитический обзор. Для электронно-печатного делопроизводства с фиксированной структурой бланка печатного документа сформулируем постановку задачи: разработать механизм идентификации, автоматизированного ввода рукописных данных и верификации соответствия электронного документа печатному аналогу с минимизацией стоимости программно-аппаратных средств, использованием универсальных расходных материалов, сокращением ручных операций и времени обработки при сохранении конфиденциальности данных.

Для решения поставленной задачи требуется использование интеллектуальных алгоритмов в области распознавания рукописного текста [2, 3], применение машиночитаемого формата представления данных [4], хеш-функции, удовлетворяющей современным условиям криптостойкости. Одним из вариантов исполнения проекта является интеграция внешнего сервиса, оказывающего услуги по защите и сертификации документов [5]. В задачах идентификации личности ведутся разработки высокоэффективных программно-аппаратных комплексов [6], основные технические идеи которых могут быть применимы для защиты документов менее высокого ранга. В качестве базового механизма защиты документа может быть хеширование графических фрагментов документа с применением специальных аппаратных средств [7, 8].

Обследование технических показателей аппаратного обеспечения учебного заведения показало сложность их применения для наложения контрольных изображений, микротекстов, водяных знаков и других артефактов для последующего их сканирования и анализа. Внешние удостоверяющие сервисы рекомендуется использовать только для документов, данные в которых не являются конфиденциальными.

С учетом выявленных технико-экономических ограничений, целесообразны использование системы хранения данных учебного заведения и разработка собственной автономной подсистемы машинного ввода и верификации документов с ее интеграцией в действующую Единую информационную систему «Электронный Университет» [9].

При выборе средств защиты электронных и печатных документов следует учитывать строгость документа, его конфиденциальность и стоимость утраты. Соблюдая баланс между затратами на разработку и обслуживание информационной системы и ее достаточной для определенных видов документов и персонала эффективностью функционала, можно добиться наилучшего технико-экономических показателей делопроизводства.

Описание используемых технологий и алгоритмов. В соответствии с поставленной задачей определен механизм машинной идентификации и верификации электронных и печатных документов. Отработка технологий и алгоритмов распознавания и проверки идентичности документов проводилась на образцах экзаменационных и зачетных ведомостей промежуточной аттестации. Подобные ведомости в сессию тиражируются сотнями экземпляров и служат основой для юридической фиксации результатов образовательного процесса в вузе. В качестве машиночитаемого формата выбран матричный двумерный код быстрого реагирования (Quick Response код) или QR-код, изобретенный японской машиностроительной корпорацией Denso-Wave для автоматизации идентификации объектов, на которых он наносится. Преимущества, достигаемые при использовании данной машиночитаемой оптической метки: высокая плотность кодируемой информации на единицу площади, легкое распознавание сканируемым оборудованием, встраивание контрольной суммы в данные препятствующей искажению распознавания. Данные могут быть или прочитаны полностью и точно или не прочитаны вовсе. Среди популярных алгоритмов хеширования данных для подготовки хеш-свертки документа выбран MD5 с наложением случайных блоков данных (соли) по определенному алгоритму, что усложняет криптоанализ и повышает криптостойкость при атаке перебором по словарю (brute force attack). При генерации документа перед выводом его на печать в БД вносится соответствующая документу JSON-структура: {"id": "идентификатор", "hash": "хеш-свертка"}. JSON – минимально нагруженный формат для представления и извлечения данных, имеющий библиотеки для различных языков программирования. Идентификатор используется для нахождения в БД электронного документа, соответствующего сканируемому, хеш-свертка используется для подтверждения подлинности электронного документа. При формировании бланка документа рассчитывается хеш-свертка от метаданных документа. При защите документа и его закрытии, хеш-свертка рассчитывается от метаданных и введенных в результате распознавания и коррекции данных документа.

Заполненный документ передается на устройство видео ввода в режиме черно-белого сканирования. В результате сканирования ненормализованная скан-копия документа (рис. 1) поступает на обработку в подсистему машинного ввода и верификации документов.

СПбГТИ(ТУ) Факультет информационных технологий и управления

Экзаменационная ведомость

Группа 411 Курс 1 Семестр 1
 ФИО преподавателя Пантелеев Владимир Александрович
 Название дисциплины Инженерная и компьютерная графика

№ п/п	ФИО студента	№ зачетной книжки	№ билета	оценка	подпись
1	Абанин Даниил Вадимович	201124	12	5 4.1	<i>de</i>
2	Власов Руслан Вадимович	200388	4	4 4.2	<i>de</i>
3	Григорьева Анастасия Анатольевна	200389	8	4 ...	<i>de</i>
4	Еремов Дмитрий Юрьевич	200390	5	3	<i>de</i>
5	Липкин Всеволод Федорович	200391	14	4	<i>de</i>
6	Лосев Анатолий Сергеевич	200392	2	3	<i>de</i>
7	Пронин Артём Андреевич	201125	10	5	<i>de</i>
8	Ситникова Елизавета Юрьевна	200270	7	5	<i>de</i>
9	Фролова Кристина Денисовна	200398	1	4	<i>de</i>
10	Щукин Алексей Сергеевич	200399	3	4 4.10	<i>de</i>

Итого 10; из них: отлично 3, хорошо 5, удовлетворительно 2,
 не удовлетворительно 0, не явилось 0.

Дата экзамена: 27 января 2021 года

Декан *de* А. А. Иванов

Преподаватель *de*

5 



Рисунок 1 – Ненормализованная скан-копия заполненного документа

На рисунке 1 красным отмечены 1 и 2 – видеометки позиционирования области распознавания, 3 – колонка таблицы для ручного ввода, 4,1, 4,2,...,4,10 – общие области для распознавания, 5 – область QR-кода. Отслеживаемые и корректируемые дефекты скан-копии: перекос, не превышающий 15 и разворот.

Нормализация полученной скан-копии документа производится по следующему алгоритму:

– проверка на разворот по отсутствию в области 4 QR-кода. При обнаружении разворота, изображение поворачивается на 180°.

– коррекция перекоса при анализе координат видеометок. Рассчитывается угол коррекции по формуле:

$$\alpha = \arctg\left(\frac{\sqrt{(x_1 - x'_1)^2 + (y_1 - y'_1)^2}}{\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}}\right),$$

где x_1, y_1 – новые координаты видеометки 1; x'_1, y'_1 – исходные координаты видеометки 1. x_2, y_2 – координаты видеометки 2. Учитывая, что при выравнивании изображения $x_1 = x_2$, а $y_1 \sim y'_1$, то формула принимает вид:

$$\alpha = \arctg\left(\frac{\sqrt{(x_2 - x'_1)^2}}{\sqrt{(y'_1 - y_2)^2}}\right).$$

Направление поворота изображения определяется при сравнении x_2 с x'_1 .

Отмечено, что выравнивание изображения эффективнее в случае наибольшей корреляции α от длины прилегающего катета. Этого можно добиться только при наибольшем расстоянии между метками 1 и 2 (рис. 2).

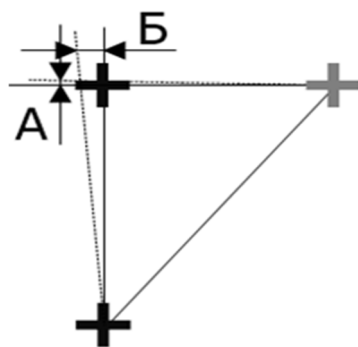


Рисунок 2 – Эффективное расположение видеометок (слева) и альтернативная видеометка (справа)

Расстояние $B > A$. Горизонтальное расположение видеометок эффективнее вертикального при портретной ориентации документа. При альбомной ориентации – наоборот.

Для поиска в общих областях распознавания 4.1, 4.2, ..., 4.10 (рис. 1) частных зон распознавания по пикселям черного цвета рукописного текста используется упрощенный метод роя частиц, показавший лучшую в сравнении с простым линейным обходом производительность (рис. 3).

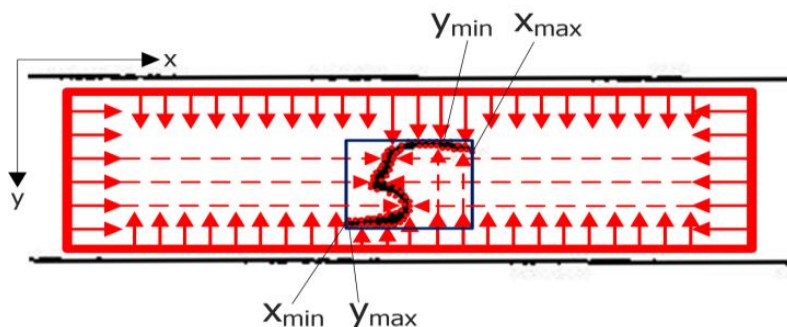


Рисунок 3 – Графическая интерпретация алгоритма поиска частной области распознавания

На рисунке 3 внешняя область – общая, внутренняя область – частная.

Производится кадрирование изображения до области, покрывающей текст для распознавания, по наименьшей и наибольшей координатам x и y из всех «осевших» на граничных пикселях черного цвета частиц.

Наиболее сложным с точки зрения реализации и настройки является механизм распознавания рукописного текста. Используется многослойная сверточная нейронная сеть, предложенная Яном Лекуном и ориентированная на эффективное распознавание образов [10]. В качестве кибернетической модели мозга используется перцептрон Розенблатта. При исследовании структуры нейросети и адаптации функций активации получено решение на базе четырех слоев – входного, двух скрытых и выходного (рис. 4).

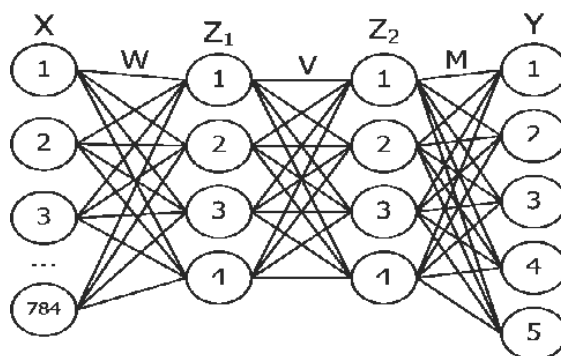


Рисунок 4 – Структура четырехслойной нейросети

На рисунке 4 X – входной слой, Z_1 и Z_2 – скрытые слои, Y – выходной слой, W , V , M – ступенчатые функции активации.

Четырехслойная нейросеть испытывалась на примере электронно-печатной экзаменационной ведомости (рис. 1). Количество входов для каждого слоя, количество слоев, алгоритмы ступенчатых функций активации W , V , M и весовые коэффициенты для каждой из них определены экспериментально.

Входами X являются пиксели масштабированной до разрешения 28x28 пикселей частной области распознавания.

Для каждого из нейронов входного слоя по координатам пикселя определена матрица весовых коэффициентов (0 или 1) по каждой из оценки (от 1 до 5). Вычисление общего веса по каждой оценке производится по цвету пикселя (рис. 5) по количеству его совпадения с эталонным по соответствующим координатам.

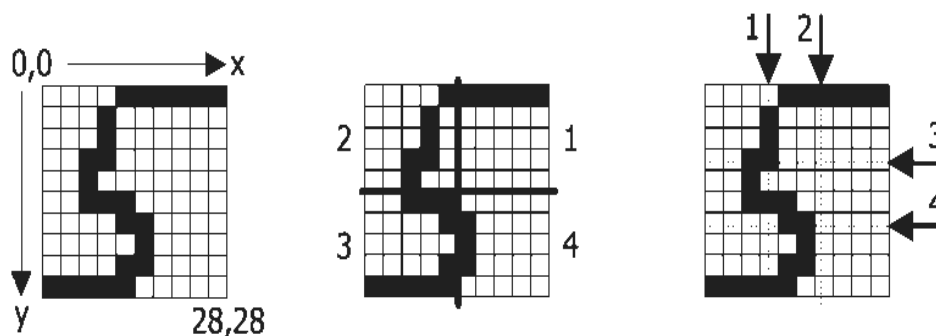


Рисунок 5 – Описание функций активации для каждого из слоев

На слое Z_1 с каждым из четырех нейронов связана структура, состоящая из четырех целых чисел (количество черных пикселей в каждом квадранте) и соответствующая им оценка. Вычисление общего веса по каждой оценке производится умножением на процент соответствия фактического количества пикселей с эталонным, установленным при обучении.

На слое Z_2 функция активации M сравнивает количество переходов белый-черный пиксель на каждом из четырех указанных направлений с эталонным количеством, установленным при обучении для каждой оценки. Общий вес складывается из процентов соответствия фактического количества переходов с эталонным.

На выходе каждый из нейронов, соответствующий оценке от 1 до 5, получает суммарный вес по нормализованной шкале от 0 до 100. Настройка доверительного порога для каждого из нейронов обеспечивает минимальную ошибку. Например, для оценки 5 установлен доверительный порог = 95. Для оценки 4 порог составляет 75. При суммарных значениях веса меньше пороговых для всех нейронов слоя Y считается отметка н/я (не явка).

Схема электронно-печатного документооборота. На рисунке 6 представлена общая схема предлагаемого электронно-печатного документооборота.

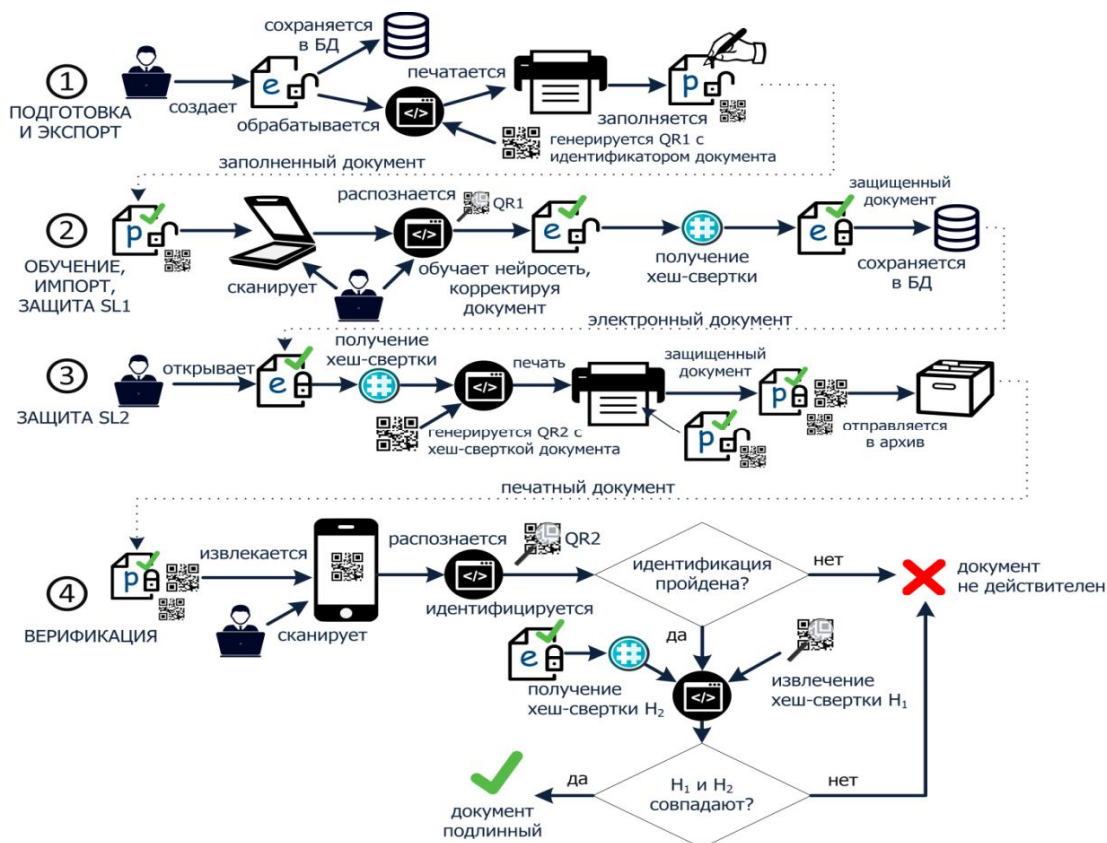


Рисунок 6 – Схема электронно-печатного документооборота

Жизненный цикл документа разбивается на 4 стадии: подготовка и печать, импорт с защитой SL1, наложение дополнительной защиты SL2 и верификация по SL2. SL1 – защита данных электронного документа, SL2 – защита от утраты соответствия электронного документа документу на бумажном носителе.

В стадии подготовки и печати для электронного документа формируется идентифицирующий QR-код (QR1) с электронным идентификатором документа и хеш-сверткой от метаданных документа. Подготовленный бланк с метаданными документа отправляется на печать для дальнейшего ручного заполнения.

В стадии импорта документа заполненный бланк сканируется, открывается изображением для визуального анализа и корректировки данных. После корректировки и сохранения данных, производится потоковое обучение нейросети с закрытием документа расчетом для него хеш-свертки от метаданных и сохраненных данных. После этого бумажный экземпляр документа может быть помещен в архив, если принято решение не накладывать на него защиту SL2.

В стадии наложения защиты SL2 генерируется верифицирующий QR-код (QR2) в виде ссылки для GET-посылки:

`https://адрес_веб-сервиса?id=идентификатор_документа&hash=хеш_свертка.`

Верифицирующий QR-код печатается на обратной (пустой) стороне документа. Документ помещается в архив.

Впоследствии, на стадии 4, при необходимости, документ извлекается из архива. На устройстве, сканирующем QR2, открывается веб-ссылка с отображением электронного экземпляра документа, данными о его подлинности и соответствии печатному аналогу.

На рисунке 7 изображена функциональная структура подсистемы машинного ввода и верификации документов ЕИС «Электронный Университет».

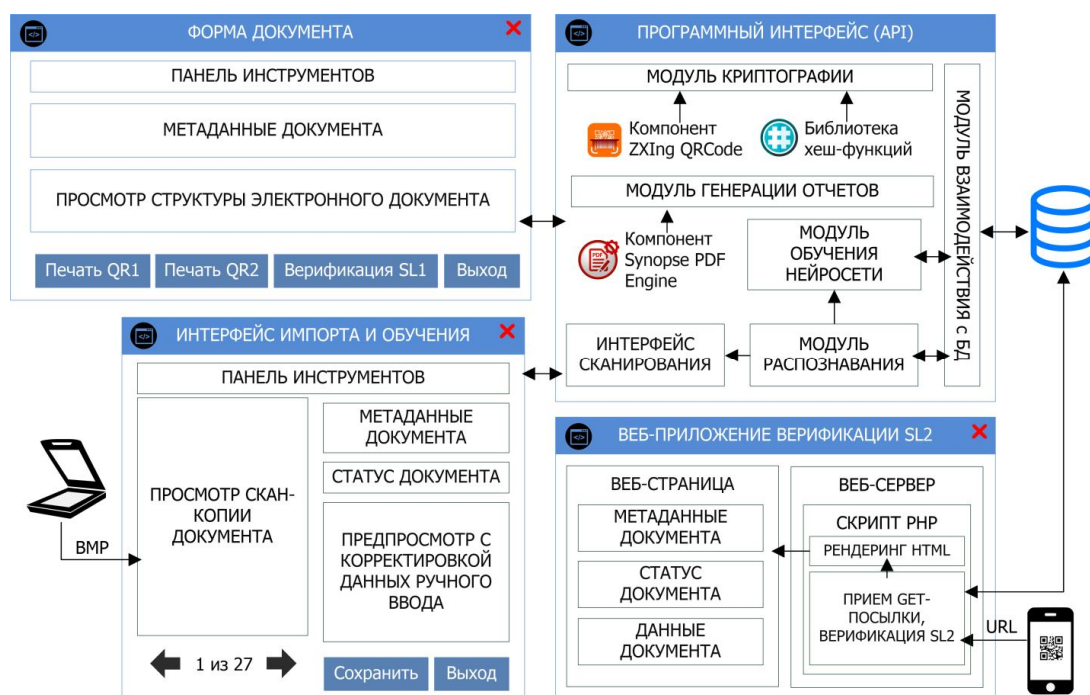


Рисунок 7 – Функциональная структура подсистемы машинного ввода и верификации документов

Графический интерфейс пользователя включает Форму документа, Интерфейс импорта и обучения и Веб-приложение верификации SL2.

Форма документа используется для подготовки и печати бланка документа с QR1 для ручного заполнения, а также для вывода на печать QR2 после ввода в электронный документ рукописных данных. Форма документа используется также для проверки подлинности электронного аналога документа (Верификация SL1).

Интерфейс импорта и обучения позволяет сканировать заполненный документ на печатном носителе, отобразить его метаданные, статус и распознанные данные ручного ввода с возможностью их корректировки. После сохранения документа выполняется ввод данных документа в БД с одновременным обучением нейросети (корректировка сохраняемых в БД весовых коэффициентов по каждому из слоев). Обучение является наиболее длительным процессом, время которого зависит от объема обрабатываемых рукописных данных.

Для ускорения обучения нейросети рассматривался вариант использования сторонней базы данных образцов рукописного написания цифр MNIST, но в итоге выбран вариант обучения нейросети на почерке работников вуза, что представляет больший научный интерес.

Веб-приложение верификации SL2 используется для приема запроса в виде URL от мобильного устройства, сканирующего QR2. При обращении к БД по идентификатору документа извлекаются метаданные и данные документа для расчета хеш-свертки, сравниваемой с полученной в URL. При равенстве хеш-сверток электронный документ получает статус «подлинный». При неравенстве хеш-сверток документ считается «не действительным». При переходе по URL открывается веб-страница с метаданными и данными документа с отображением его статуса. Веб-страница используется для визуального сравнения и анализа данных «не действительного» электронного документа.

В программном интерфейсе подсистемы машинного ввода и верификации документа используются компоненты для генерации и распознавания штрих-кодов «ZXing QRCode», библиотека с реализацией алгоритмов хеш-функций, в том числе MD5, компонент для рендеринга и вывода документов формата PDF Synopse «PDF Engine», модуль с реализацией алгоритма обучения нейросети с модулем распознавания рукописного текста. Интерфейс сканирования служит «обвязкой» между графическим Интерфейсом импорта и обучения и программным интерфейсом. Его назначение – обращение к выбранному пользователем сканирующему устройству по стандартному протоколу TWAIN с захватом изображения и передачей его для просмотра в компонент TImage Интерфейса импорта и обучения.

Заключение. Таким образом, предложенная методика и ее практическая реализация позволяют без существенного изменения традиционного документооборота выстроить единую систему электронно-печатного делопроизводства организации. При этом время формирования электронного документа из его печатного аналога сокращается в среднем в три раза. Обеспечивается достаточный уровень защищенности электронных документов и их оперативная верификация по аналогам на бумажном носителе.

Заложенные в систему функции самообучения, а также автоматической маркировки подлинности их электронных аналогов, не требуют проведения дополнительных организационных мероприятий, применения специального оборудования и специфических бумажных носителей.

Используемая технология автоматизированного перевода документов на бумажных носителях в электронный вид обладает известной гибкостью. В зависимости от важности документа и от поставленных задач изменяется, например, степень его защиты с SL2 на SL1.

В перспективе планируется создание отдельного сервиса приложений с API для распознавания рукописного текста на базе Python, библиотеки Keras и высокопроизводительной векторной СУБД.

Тестирование полученного программного решения доказало его применимость для многотиражных документов строгой отчетности фиксированной структуры с рукописным посимвольным заполнением.

Библиографический список

1. Провоторский А. О. Методы и средства защиты бумажных документов от подделки при минимизации стоимости листа / А. О. Провоторский // Научный Альманах. – 2016. – № 4–3 (18). – С. 156–158. – DOI 10.17117/na.2016.04.03.156.
2. Jia Y. A lightweight handwriting recognition system based on an improved convolutional neural network / Y. Jia, Y. Zhao, Y. Zhang, S. Fan // The ACM International Conference Proceeding Series. – 2020. – P. 30–35. – DOI 10.1145/3436369.3436463.
3. Behera R. K. A smart document converter: Conversion of handwritten text document to computerized text document / R. K. Behera, B. Padhi // Studies in Computational Intelligence. – 2021. – № 907. – P. 153–166. – DOI 10.1007/978-3-030-50641-4_9123.
4. Li C. M., Hu P., Lau W. C. AuthPaper – Protecting paper-based documents/credentials using authenticated 2D barcodes / C. M. Li, P. Hu, W. C. Lau // MobiSys : proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services. – 2014. – 348 p. – DOI 10.1145/2594368.2601468.
5. IDX: официальный сайт. – ООО «Системы управления идентификацией», 2021. – Режим доступа: <https://iidx.ru>, свободный. – Заглавие с экрана. – Яз. рус. (дата обращения: 12.04.2021).
6. Azhmukhamedov I. Identification of personality based on electronic documents with increased security level / I. Azhmukhamedov, N. Poletayev, A. Stanishevskaya // Caspian Journal: Control and High Technologies, 2019. – № 2 (46). – P. 170–179.
7. Ahmad F. Paper Document Authentication Using Print-Scan Resistant Image Hashing and Public-Key Cryptography (2019) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) / F. Ahmad, L.-M. Cheng // 11611 LNCS. – 2019. – P. 157–165. – DOI 10.1007/978-3-030-24907-6_13.
8. Патент № RU 157246 U1 Российская Федерация. Система защиты от подделок многотиражных документов строгой отчетности : № 2014147705/08 : заявл. 26.11.2014 : опубл. 27.11.2015 / Д. А. Блудов, В. Н. Богданов, П. С. Вихлянцева ; заявитель Федеральное государственное унитарное предприятие «ЦентрИнформ».
9. Свидетельство о государственной регистрации программы для ЭВМ № 2018618919 Российская Федерация. Единая информационная система «Электронный Университет» : № 2018615462 : заявл. 28.05.2018 : опубл. 23.07.2018 / Д. Н. Петров, М. Ю. Васильев, И. В. Бабичев ; заявитель Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный технологический институт (технический университет)» (СПбГТИ (ТУ)).
10. Andreeva E. I. Document Recognition Method Based on Convolutional Neural Network Invariant to 180 Degree Rotation Angle / E. I. Andreeva, V. V. Arlazarov, A. V. Gayer et al. // Journal of Information Technologies and Computing Systems. – 2019. – № 4. – P. 87–93. – DOI 10.14357/20718632190408.

References

1. Provotorskiy, A. Metody i sredstva zashchity bumazhnykh dokumentov ot poddelki pri minimizatsii stoimosti lista [Methods and means of protection of paper documents from a fake at minimization of cost of a leaf]. *Nauchnyy Almanakh* [Science Almanac], 2016, no. 4–3 (18), pp. 156–158. DOI 10.17117/na.2016.04.03.156.
2. Jia Y., Zhao Y., Zhang Y., Fan S. A lightweight handwriting recognition system based on an improved convolutional neural network. *The ACM International Conference Proceeding Series*, 2020, pp. 30–35. DOI 10.1145/3436369.3436463.
3. Behera R. K., Padhi B. A. Smart document converter: Conversion of handwritten text document to computerized text document. *Studies in Computational Intelligence*, 2021, no. 907, pp. 153–166. DOI 10.1007/978-3-030-50641-4_9.

4. Li C. M., Hu P., Lau W. C. AuthPaper – Protecting paper-based documents/credentials using authenticated 2D barcodes. *MobiSys. Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, 2014. 348 p. DOI 10.1145/2594368.2601468.
5. *IDX: ofitsialnyy sayt – OOO «Sistemy upravleniya identifikatsiyey»* [IDX: official site – OOO «Identity management systems»], 2021. Available at: <https://iidx.ru> (accessed 12.04.2021).
6. Azhmukhamedov I, Poletayev N. Stanishevskaya A. Identification of personality based on electronic documents with increased security level. *Caspian Journal: Control and High Technologies*, 2019, no. 2 (46), pp. 170–179.
7. Ahmad F., Cheng L.-M. Paper Document Authentication Using Print-Scan Resistant Image Hashing and Public-Key Cryptography. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 11611 LNCS*, 2019, pp. 157–165. DOI 10.1007/978-3-030-24907-6_13.
8. Bludov D. A., Bogdanov V. N., Vikhlyantsev P. S. *Sistema zashchity ot poddelok mnogotiraznykh dokumentov strogoj otchetnosti* [The System of protection against counterfeiting of high-circulation documents of strict accountability]. Patent, FSUE «CentrInform», no RU 157246 U1, 2015.
9. Petrov D. N., Vasilyev M. U., Babichev I. V. *Edinaya informatsionnaya sistema «Elektronnyy Universitet». Svidetelstvo ob ofitsialnoy registratsii programm dlya EVM* [Unified information system "Electronic University", Certificate of state registration of a computer program], Saint-Petersburg State Technological Institute (Technical University), 2018, no. 2018618919.
10. Andreeva E. I., Arlazarov V. V., Gayer A. V. et al. Document Recognition Method Based on Convolutional Neural Network Invariant to 180 Degree Rotation Angle. *Journal of Information Technologies and Computing Systems*, 2019, no. 4, pp. 87–93. DOI 10.14357/20718632190408.