

(19)



**Евразийское
патентное
ведомство**

(11) **039140**(13) **B1**(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**

(45) Дата публикации и выдачи патента
2021.12.09

(21) Номер заявки
201992843

(22) Дата подачи заявки
2019.12.26

(51) Int. Cl. **G06Q 50/18** (2012.01)
G06N 20/00 (2019.01)
G06N 3/08 (2006.01)

(54) СПОСОБ И СИСТЕМА АВТОМАТИЧЕСКОГО ПРИНЯТИЯ ПРАВОВОГО РЕШЕНИЯ

(31) **2019139543**

(32) **2019.12.04**

(33) **RU**

(43) **2021.06.30**

(71)(73) Заявитель и патентовладелец:
**ПУБЛИЧНОЕ АКЦИОНЕРНОЕ
ОБЩЕСТВО "СБЕРБАНК
РОССИИ" (ПАО СБЕРБАНК) (RU)**

(72) Изобретатель:
**Карпец Михаил Валерьевич, Накипов
Искандер Наилевич, Денисов Илья
Вячеславович, Емельянов Ярослав
Игоревич, Волкова Ольга Сергеевна,
Новиков Михаил Юрьевич, Кузнецов**

**Максим Викторович, Буракова
Марина Валериевна, Крылова Дарья
Андреевна, Клыков Глеб Игоревич,
Шульга Сергей Александрович (RU)**

(74) Представитель:
Герасин Б.В. (RU)

(56) CN-A-106934483
MASHA MEDVEDEVA et al., "Using
Machine Learning to Predict Decisions of the
European Court of Human Rights", Artificial
Intelligence and Law, 26.06.2019, 30 л., найдено в
Интернет: <https://link.springer.com/content/pdf/10.1007/s10506-019-09255-y.pdf>, реферат, разделы
1, 3, 4
US-A1-20190266196
US-A1-20170116519

(57) Изобретение позволяет оптимизировать юридическую деятельность под заданные условия с помощью автоматизации и обеспечения комплексного решения задач принятия правового решения по заранее заданному запросу на основании любых документов клиента, определяя степень их относимости к рассматриваемому вопросу, синтезируя их актуальные версии (при наличии изменений), выявляя наличие в них имеющих значение для решения поставленной задачи юридических фактов, анализируя выявленные юридические факты (при их наличии), выявляя и обрабатывая при этом недостатки в заполнении данных документов. Применение данного изобретения не требует адаптации под него автоматизируемых процессов и применяемых подходов, в связи с чем получаемые на вход документы не требуют специального отбора, исправления ошибок или иной предобработки, а формируемые по итогам анализа заключения идентичны формируемым юристами или иными специалистами с помощью сформированных глоссария и правил выполнения требуемой обработки. Изобретение разработано для принятия решений в области права, но его имплементация в другие области техники и знаний не ограничена и может масштабироваться под заданные условия его применения. Используя данное изобретение, пользователь может обеспечить автоматизацию процесса принятия решений в соответствующей обозначенной им области знаний с учетом логики принятия решений, необходимых для этого параметров и требований, имеющих в автоматизируемом им процессе и задаваемых пользователем самостоятельно. При этом реализация изобретения не требует от пользователя наличия специальных знаний в области программирования. Техническим результатом от реализации изобретения является обеспечение автоматизированной проверки документов для принятия правового решения на основании поступающего запроса с помощью анализа и выявления соответствия имеющихся в них юридических фактов на основании знаний предметной области техники с помощью алгоритмов машинного обучения. Заявленный результат достигается за счет применения системы интеллектуального анализа документов, подлежащих правовой проверке в рамках поступающего запроса пользователя, которая осуществляет автоматизированную обработку информации с помощью алгоритмов машинного обучения, обеспечивающих функционал по выявлению и проверке юридических фактов в таких документах, их сопоставлению и дальнейшему принятию правового решения, отражаемого в юридическом заключении.

B1**039140****039140 B1**

Область техники

Изобретение в общем относится к вычислительной области техники, а в частности к способам и системам автоматического принятия правового решения на основании обработки документов.

Уровень техники

В настоящее время работа юридического департамента крупной компании сопряжена не только с анализом и обработкой больших массивов данных, но и с необходимостью внедрения комплексных систем автоматизации учета, контроля и управления данными, например, о контрагентах.

Общее ускорение бизнес-процессов приводит к тому, что юристам необходимо подготавливать документы и принимать решения в кратчайшие сроки, обрабатывая при этом большие массивы данных. В результате этого становится невозможно эффективно выполнять задачи, работая устаревшими методами, ведя расчеты на бумаге или в стандартных программах, контролируя ход дел по обычным календарям и спискам. Ручная обработка и проверка данных о контрагентах на основании большого объема информации может привести к катастрофическим ошибкам для бизнеса.

Из уровня техники известны различные технические решения в области подходов по автоматизации анализа данных, относящихся к юридическим текстам или иному виду нормативно-правовых документов. Например, в патенте США US 6772149 (правообладатель: LexisNexis Group, дата публикации: 03.08.2004) раскрывается подход по машинному анализу текстов судебных актов, с помощью выявления юридических фактов, содержащихся в теле того или иного документа с их последующей классификацией, которая впоследствии может применяться для обучения модели машинного обучения. Из другой патентной заявки США US 20190266196 (правообладатель: Confidentiality Corp, дата публикации: 29.08.2019) известно решение по анализу документов, содержащих, например, юридические тексты, для целей выявления связанных документов по тому или иному типу связей (фактов, атрибутов и т.п.) и выявлению схожих документов или их частей.

Известные из уровня техники решения тем или иным образом позволяют оптимизировать юридическую деятельность под заданные условия, передав часть работы на автоматизацию, однако не обеспечивают комплексного решения задачи по автоматизации всего процесса принятия правового решения, в частности, принятие правового решения по заранее заданному запросу на основании любых документов клиента, определяя степень их относимости к рассматриваемому вопросу, выявляя наличие в них имеющих значение для решения поставленной задачи юридических фактов, анализируя выявленные юридические факты (при их наличии), выявляя и обрабатывая при этом недостатки в заполнении данных документов.

Сущность изобретения

Заявленное изобретение предлагает новый подход в области алгоритмизации проведения правовой экспертизы и формирования юридических заключений по заданной теме.

Решаемой технической проблемой или технической задачей является создание автоматизированного способа принятия правовых решений и системы формирования юридических заключений.

Основным техническим результатом, достигающимся при решении вышеуказанной проблемы, является обеспечение автоматизированной проверки документов для принятия правового решения на основании поступающего запроса с помощью анализа и выявления соответствия имеющихся в них юридических фактов на основании знаний предметной области техники с помощью алгоритмов машинного обучения.

Дополнительным техническим результатом, достигающимся при решении вышеуказанной проблемы, является повышение скорости принятия правовых решений, а также анализ документов на соответствие требованиям нормативно-правовых актов и внутренней нормативной документации (далее - ВНД) компании.

Заявленные результаты достигаются за счет осуществления компьютерно-реализуемого способа автоматического принятия правового решения, выполняемого с помощью по меньшей мере одного процессора и содержащего этапы, на которых

получают запрос на формирование правового решения;

получают по меньшей мере один документ, соответствующий упомянутому запросу;

обрабатывают полученный документ на предмет выявления в документе по меньшей мере одной сущности, относящейся к запросу правового решения, и осуществляют ее извлечение из документа с помощью алгоритма машинного обучения;

осуществляют юридический анализ упомянутой по меньшей мере одной извлеченной сущности с помощью набора правил, сформированных с помощью предметно-ориентированного языка;

формируют правовое решение на основании проведенного юридического анализа упомянутых извлеченных сущностей упомянутого по меньшей мере одного документа.

В одном из частных вариантов осуществления способа на этапе обработки документа осуществляется распознавание печатного текста документа, при котором выполняется автоматическое распознавание изображения документов и извлечение текстовых данных из них.

В другом частном варианте осуществления способа на основании извлеченных текстовых данных формируется новый документ.

В другом частном варианте осуществления способа выполняется определение типа документа с помощью сравнения извлеченных текстовых данных с шаблоном документа.

В другом частном варианте осуществления способа при получении нескольких документов осуществляется их обработка на предмет объединения в рамках полученного запроса формирования правового решения.

В другом частном варианте осуществления способа выполняется сравнение изображения печатного текста документа с новым документом, сформированным по итогам извлечения текстовых данных, сохраненным в новом документе, образованном в результате распознавания.

В другом частном варианте осуществления способа при наличии расхождения между документами в результате сравнения, в новом документе выполняют исправления ошибок распознавания изображения документа.

В другом частном варианте осуществления способа дополнительно выполняют верификацию по меньшей мере одной извлеченной сущности из документа. В другом частном варианте осуществления способа выполняется анализ смыслового соответствия по меньшей мере одной извлеченной сущности участку текста документа, из которого было осуществлено упомянутое извлечение.

В другом частном варианте осуществления способа выполняется корректировка по меньшей мере одной извлеченной сущности, причем такая корректировка включает в себя по меньшей мере одно из: исправление орфографических ошибок, пунктуации, технических ошибок, смысловых ошибок.

В другом частном варианте осуществления способа выполняется фильтрация документов, по меньшей мере, утративших юридическую силу.

В другом частном варианте осуществления способа на этапе выполняется кросс-проверка юридических сущностей, при которой осуществляется сравнение юридических фактов, содержащихся в документах, отобранных для формирования юридического заключения.

Заявленное изобретение также осуществляется с помощью устройства автоматического принятия правового решения, и данное устройство содержит по меньшей мере один процессор и по меньшей мере одно средство памяти, соединенное с процессором и содержащее машиночитаемые инструкции, которые при их исполнении процессором обеспечивают выполнение вышеуказанного способа.

В другом предпочтительном варианте осуществления заявленного технического решения представлена система автоматического принятия правового решения, содержащая

систему интеллектуального распознавания документов, включающую модуль обработки, выполняющий

получение запроса на формирование правового решения;

получение по меньшей мере одного документа, связанного с упомянутым запросом;

модуль извлечения сущности, обеспечивающий выявление в документе по меньшей мере одной сущности, относящейся к запросу правового решения, и ее извлечение из документа с помощью алгоритма машинного обучения; систему автоматизированного принятия решений, обеспечивающую

выполнение юридического анализа упомянутой по меньшей мере одной извлеченной сущности с помощью набора правил, сформированных с помощью предметно-ориентированного языка;

формирование правового решения на основании проведенного юридического анализа упомянутых извлеченных сущностей упомянутого по меньшей мере одного документа.

В одном из частных примеров реализации система дополнительно содержит модуль верификации извлеченных сущностей, выполняющий проверку правильности извлечения атрибутов.

В другом частном примере реализации системы модуль верификации выполняет анализ смыслового соответствия по меньшей мере одной извлеченной сущности участку текста документа, из которого было осуществлено упомянутое извлечение.

В другом частном примере реализации системы модуль верификации выполняет корректировку по меньшей мере одной извлеченной сущности, причем такая корректировка включает в себя по меньшей мере одно из: исправление орфографических ошибок, пунктуации, технических ошибок, смысловых ошибок.

Краткое описание чертежей

Признаки и преимущества настоящего изобретения станут очевидными из приводимого ниже подробного описания изобретения и прилагаемых чертежей, на которых:

фиг. 1 иллюстрирует блок-схему выполнения заявленного способа;

фиг. 2 иллюстрирует фрагмент с текстом документа, из которого производится извлечение атрибутов;

фиг. 3 иллюстрирует пример результата извлечения сущности;

фиг. 4 иллюстрирует пример разметки данных;

фиг. 5 иллюстрирует пример дочерних документов;

фиг. 6 иллюстрирует пример сообщения о проверке сущностей;

фиг. 7 иллюстрирует пример сформированного правового заключения;

фиг. 8 иллюстрирует пример работы заявленного решения для целей предоставления услуг клиенту;

фиг. 9 иллюстрирует блок-схему процесса обработки запроса на осуществление проверки докумен-

тов клиента;

фиг. 10 иллюстрирует общий вид вычислительного устройства для реализации заявленного способа.

Подробное описание изобретения

Заявленное изобретение может применяться для формирования правовых решений различного профиля и характера, исходя из требуемого запроса. Например, проверка правоспособности контрагента, в ходе которой будут в автоматическом режиме проанализированы все его документы, в том числе на предмет выявления единоличного исполнительного органа и подтверждения сроков его полномочий, проверка и подтверждение полномочий органов управления, выявление ограничений на совершение определенных сделок и т.п. Основной особенностью решения является обеспечение в автоматизированном режиме интеллектуального анализа входящих документов для генерирования правового решения. Итоги проведенной проверки автоматически вносятся в правовое решение, формируя итоговое решение по вопросу. При отсутствии документов или неактуальности имеющихся будет сформирован соответствующий список для запроса у клиента с обоснованием причин такого запроса. Данное изобретение может быть реализовано на компьютере, в виде автоматизированной информационной системы (АИС) или машиночитаемого носителя, содержащего инструкции, например, программные, для выполнения вышеупомянутого способа.

Изобретение может быть реализовано в виде распределенной компьютерной системы, которая может быть установлена на централизованном сервере (наборе серверов). Доступ пользователей к системе возможен как из сети Интернет, так и из внутренней сети предприятия/организации посредством мобильного устройства связи, на котором установлено программное обеспечение с соответствующим графическим интерфейсом пользователя, или персонального компьютера с доступом к веб-версии системы с соответствующим графическим интерфейсом пользователя. Ниже будут описаны термины и понятия, необходимые для реализации настоящего технического решения.

В данном решении под системой подразумевается компьютерная система, ЭВМ (электронно-вычислительная машина), ЧПУ (числовое программное управление), ПЛК (программируемый логический контроллер), компьютеризированные системы управления и любые другие устройства, способные выполнять заданную, четко определённую последовательность вычислительных операций (действий, инструкций). Под устройством обработки команд подразумевается электронный блок либо интегральная схема (микропроцессор), исполняющая машинные инструкции (программы). Устройство обработки команд считывает и выполняет машинные инструкции (программы) с одного или более устройств хранения данных. В роли устройства хранения данных могут выступать, но, не ограничиваясь, жесткие диски (HDD), флэш-память, ПЗУ (постоянное запоминающее устройство), твердотельные накопители (SSD), оптические приводы.

Программа - последовательность инструкций, предназначенных для исполнения устройством управления вычислительной машины или устройством обработки команд. Способ автоматического принятия решения для юриста показан на фиг. 1 (100) в виде блок-схемы и далее будет подробно поэтапно раскрыт.

На первом этапе (110) выполняется получение запроса на формирование правового решения. В запросе содержится указание на специфику требуемой проверки, в ходе которой предоставляется один или несколько документов (120). В качестве документов, полученных на этапе (120) могут представляться отсканированные копии оригиналов документов, или цифровые формы документов. Данные документы могут получать из внутренних хранилищ данных организации или из внешних источников информации. Например, если в запросе содержится требование проведения проверки юридического лица с целью выдачи кредита, то среди предоставленных документов могут присутствовать уставы, протоколы и т.д. Запрос на формирование может не ограничиваться только одной областью знаний, относящейся к области юриспруденции, и может также применяться для иной области, в которой также возможно формирование решение (заключение) по тому или иному вопросу.

Полученный документ или ряд документов в ходе обработки запроса на этапе (120) передается в систему интеллектуального распознавания документов (далее - СИРД). Система интеллектуального распознавания документов (СИРД) состоит из модуля обработки, модуля извлечения атрибутов и модуля верификации извлеченных атрибутов. Если документ ранее уже проходил все модули СИРД, и в досье клиента был сформирован соответствующий дочерний документ, то данный документ повторно не проходит модули СИРД - дочерний документ сразу передается в систему автоматического принятия решения. Описание дочернего документа будет дано далее в настоящем описании.

Набор документов может храниться в форматах XML, XLS, DBF, CSV, TXT и изображений TIFF, JPEG, JPEG 2000, PDF, PDF/A, PCX, BMP, PNG, или иных поддерживаемых форматах, не ограничиваясь.

Далее на этапе (130) выполняется обработка полученного запроса и соответствующих ему одного или нескольких документов для выявления и извлечения сущностей, содержащих атрибуты, относящиеся к юридическому факту, связанному с запросом на формирование правового заключения.

В частности, при исполнении этапа (130) выполняется распознавание и типизация документов, верификация распознанного текста, объединение документов (при необходимости). Распознавание доку-

ментов осуществляется на базе технологии OCR (англ. Optical Character Recognition), которая позволяет автоматически перевести изображение печатного текста документа в текстовые данные (набор текстовых символов). Технология представляет собой комплекс действий, включающих обработку изображений, детектирование в них текста и его распознавание, и может выполняться с помощью известных алгоритмов и программно-аппаратных подходов.

Действия по обработке изображения направлены на улучшение качества документа и повышение качества детектирования текста. Для детекции текста используется ИНС, архитектурно представляющая собой CNN (сверточная нейросеть), обученную на наборе из 5000 документов, содержащих размеченные блоки текста. Для распознавания текста используется ИНС, основанная на архитектуре CRNN (сверточная рекуррентная нейросеть). Данная нейросеть обучена на наборе синтетических (сгенерированных) данных, представляющих собой образцы текста, размещенные на нескольких типовых фонах, которые наиболее часто встречаются в документах. Каждый образец представляет собой пару элементов: текст на фоне в формате изображения и его расшифровка в текстовом виде. Для обучения было использовано около 100000 наиболее частотных слов русского языка, на основе которых было сгенерировано больше 10 млн изображений. Одновременно может осуществляться типизация распознаваемого документа. Для определения типа документа, модуль обработки соотносит набор текстовых символов документа с шаблонной структурой, характеризующей определенный тип документа. Процесс соотнесения представляет собой поиск в распознанном текстовом слое характерных словесных конструкций, присущих определенному шаблону. Шаблон (шаблонная структура) создается на основании алгоритмов, которые анализируют вариативность структуризации различных типов документов. Данные алгоритмы устанавливают закономерности в структуре документов и формируют на основании этого шаблон. Например, шаблонной структурой в уставе является наличие разделов с характерным наименованием ("Общие положения", "Уставной капитал" и т.д.), упорядоченность следования пунктов, наличие титульного листа с указанием слова "Устав" и т.д.

Опционально может выполняться этап верификации (131) с помощью модуля верификации. Полнотекстовая верификация производится путем сравнения изображения печатного текста документа с набором текстовых символов, сохраненным в новом документе, образованном в результате распознавания. При наличии расхождений между изображением печатного текста и набором текстовых символов в новом документе, модуль обработки производит исправление ошибок распознавания, в частности выполняется изменение и замена некорректно распознанных текстовых символов на достоверные символы.

По результатам распознавания и типизации набор текстовых символов по отдельному документу, сохраняется во вновь созданном документе. Исходные документы используются в формате html, doc, txt, pdf и иных форматах, не ограничиваясь. Далее модуль обработки определяет необходимость объединения типизированных документов. Объединение производится, если на этапе (120) было предоставлено два и более документа (набор документов), относящихся к обработке полученного запроса на этапе (110), при наличии в наборе документов основного документа и другого документа, содержащего изменения текста основного документа. Например, при выявлении в пакете документов Устава (основной документ) и Изменения к уставу (изменяющий документ) система направляет их на объединение с целью формирования финальной версии соответствующего документа.

Процесс объединения документов состоит из следующих шагов: определение структуры объединяемых документов, определение объема и типа изменений, применение выявленных изменений (непосредственно объединение).

Определение структуры основного (изменяемого) документа заключается в разделении его на блоки информации в соответствии с примененной в документе нумерацией. Например, блоками основного документа могут являться пункты, статьи или параграфы. Определение структуры изменяющего документа заключается в выявлении текста всех правок, которые необходимо внести в изменяемый документ. Под правкой понимается выраженное на естественном языке указание на изменение текста основного документа. Например, правкой будет являться следующий текст в Изменениях к уставу: "Пункт 38 раздела 2 Устава Общества изложить в следующей редакции: "Размер уставного капитала Общества составляет 150000 рублей".

Каждая правка подразделяется на элементы: заголовок правки (например, "пункт 38 раздела 2 Устава Общества изложить в следующей редакции") и содержание правки (например, "размер уставного капитала Общества составляет 150000 рублей"). Содержание правки является опциональным полем. Например, правки на удаление текста представляются только заголовком.

Затем определяется тип заголовка правки. Заголовок подразделяется на типы: изменение блока, добавление, удаление, изменение слов и фраз по всему тексту. После определения типа заголовка правки осуществляется извлечение адреса блока из заголовка правки. Под адресом блока понимается конкретный номер пункта, номер статьи абзаца или непосредственное слово, или фраза, подлежащая изменению (например, "генеральный директор" заменить на "директор"). Затем извлеченная информация (адрес блока, правка, тип заголовка), которая по сути является входными данными для объединения документов, проходит обработку модулем обработки для исполнения операций по объединению документов. Модуль обработки выполняет операцию по объединению документов - действие по физической замене исходно-

го текста на текст правки, добавлению или удалению на основании входных данных. В результате данной операции образуется новый документ, в котором к исходному документу применены все операции правок (внесены изменения).

Далее с помощью модуля извлечения сущности инициируется процесс извлечения атрибутов из обработанных документов. Под атрибутом в данном техническом решении понимается единица информации, формирующая сущность, относящуюся к юридическому факту. Например, наименование субъекта и населенного пункта, зафиксированные в Уставе в качестве местонахождения организации (см. фиг. 2), формируют два соответствующих атрибута (фиг. 3) - "Субъект РФ" и "НасПункт". Под сущностью в данном техническом решении понимается набор атрибутов, объединенных в одну группу по какому-либо признаку, относящихся к юридическому факту. На основе примера, приведенного выше, атрибуты, указывающие на наименование субъекта и населенного пункта, формируют сущность "местонахождение". На фиг. 3 - "МН".

В приведенном примере из Устава, объединенного с Изменениями к нему, были извлечены следующие данные по местонахождению организации (сущность "МН"): субъект - "Ярославская" (атрибут "СубъектРФ"), населенный пункт - "Переславль-Залесский" (атрибут "НасПункт").

Состав извлекаемых атрибутов и сущностей формируется пользователем применительно к каждому документу в отдельности и представляет собой Глоссарий атрибутов. Глоссарий - отражает область знаний, в которой применяется тот или иной документ. Таким образом, его формирование осуществляется специалистом в соответствующей предметной области самостоятельно - без привлечения специалистов, обладающих специальными знаниями и навыками в области программирования. Извлечение атрибутов производится автоматически путем информационного поиска сущности и выделения из ее состава юридически значимой информации (атрибутов) с помощью алгоритмов машинного обучения. Разработка алгоритмов машинного обучения включает в себя следующие основные этапы, не ограничиваясь: определение источника данных, формирование на их основе обучающей выборки, подбор оптимальных методов машинного обучения на полученной обучающей выборке, обучение. Описание этапов приведено ниже.

Машинное обучение проводится на клиентских данных, а также данных, раскрытых в рамках корпоративного законодательства. Под раскрытием информации в рамках исполнения требований корпоративного законодательства следует понимать установленные законными и подзаконными актами нормы права, обеспечивающие доступность информации всем заинтересованным лицам, независимо от целей получения данной информации в соответствии с процедурой, гарантирующей ее нахождение и получение.

Для формирования обучающей выборки для алгоритмов машинного обучения, применяемых в настоящем техническом решении, используются выборки с различными типами документов для выявления в них различных юридических фактов. Количество документов в выборках составляет от 100 до 5000 документов (например, различные типы уставов, договоров и т.п.). После отбора документов инициируется их распознавание (описание технологий приведено выше) и разметка.

Разметка представляет собой выделение в тексте юридических фактов с указанием, к какой юридической сущности и атрибуту он относится. Пример представлен на фиг. 4. Разметка может выполняться с помощью обученной модели машинного обучения при извлечении юридических фактов из текста.

Обучение производится в рамках решения следующих задач: классификация, регрессия, написание детерминированных алгоритмов. В настоящем решении используется ансамбль из 16 моделей машинного обучения, которые обучаются от 3 до 18 эпох в зависимости от момента достижения требуемого показателя точности выявления и классификации данных. Модели, представленные в ансамбле, являются архитектурами следующих видов: CNN (сверточная сеть), RNN (рекуррентная сеть), которые могут включать в себя специальные слои LSTM (долгая краткосрочная память), механизмы self-attention (контроль внимания) и др. Также используются доработанные модели представления текста: w2v, elmo и др.

Результатом прохождения этапа (130) является формирование файла в формате xml, который может также при необходимости передаваться в модуль верификации извлеченных атрибутов на этапе (131).

Как представлено на фиг. 3 файл содержит в том числе следующую информацию: тип документа, извлеченные атрибуты и сущности. Каждой единице извлеченной информации в файле соответствует путь следующего вида, не ограничиваясь: ТипДокумента\Сущность\Атрибут (фиг. 3).

При выполнении этапа (131) модуль верификации производит проверку правильности извлечения атрибутов, полученных на этапе (130), с целью передачи их в систему автоматического принятия решения.

Проверка юридического факта, содержащегося в извлеченном атрибуте, производится путем поиска данного факта в наборе текстовых символов в соответствующем документе и последующей проверки правильности произведенного извлечения на основании контекстного поиска (смыслового соответствия извлеченного атрибута участку текста, где он был найден). Например, данная проверка позволяет выявлять ситуации, когда вместо цифры, характеризующей уставной капитал, в соответствующий атрибут попала цифра, отвечающая за количество членов совета директоров. Если искомым атрибутом отсутствует или не соответствует набору текстовых символов, модуль верификации извлеченных атрибутов вносит

исправления в файл, удаляя или корректируя рассматриваемый атрибут.

Результатом прохождения этапа верификации (131) извлеченных атрибутов является сохранение итогового файла, сформированного на этапе (130) с выполненными корректировками (при необходимости) в хранилище документов.

Хранилище документов содержит документы, представленные в ходе обработки запроса (110) на этапе (120) или полученные из смежных систем для выполнения формирования правового решения. В хранилище хранятся как основные, так и дочерние (связанные) документы.

Под дочерним документом в данном техническом решении понимается документ, сформированный по итогам прохождения этапа извлечения атрибутов из документа, представленного клиентом, и этапа их последующей верификации. Например, к Протоколу об избрании единоличного исполнительного органа, размещенному в досье клиента в формате pdf, после прохождения этапа верификации извлеченных из него атрибутов будет сохранен дочерний документ в формате xml, содержащий извлеченные из него сущности и атрибуты (фиг. 5).

Далее система интеллектуального распознавания документов (СИРД) передает сохраненный файл, сформированный на этапе (130) и скорректированный при необходимости на этапе (131), в систему автоматического принятия решения, которая содержит хранилище набора правил для формирования правового решения и обеспечивает последующую обработку для формирования юридического анализа.

На этапе (140) выполняется применение сформированного набора правил для юридического анализа извлеченных сущностей из обработанных документов. На данном этапе система автоматического принятия решений автоматически преобразовывает в скрипт на языке Python каждое правило, ранее подготовленное и написанное на предметно-ориентированном языке, которые хранятся в специализированном хранилище (141), например, базе данных.

Созданный для решения данной задачи предметно-ориентированный язык позволяет формировать необходимые наборы правил пользователю, не обладающему специальными знаниями и навыками в области программирования.

Правило подразумевает под собой набор действий, аналогичный логике и порядку действий, производимых человеком (юристом) во время проведения проверки, анализа юридических фактов и принятия решения. Каждое правило представляет собой анализ извлеченных атрибутов и сущностей посредством использования операторов кода. Анализируемые атрибуты типизируются в зависимости от вида информации, которую они содержат, и специфики проводимой проверки. Примерами типов атрибутов могут служить: строка, число, дата и т.д. Типизация позволяет осуществлять специфические для соответствующего типа атрибутов операции. Например, к атрибутам, типизированным как "число", возможно применение математических операторов. Под оператором кода в данном техническом решении понимается заранее предустановленный формализованный список команд для реализации возможности написания алгоритмов с использованием правил логики. Команды представляют собой набор логических конструкций и применяемых математических операторов. Примерами операторов кода могут выступать, но, не ограничиваясь, конструкции типа: "если ... то", "и", "содержит" и т.д.

Результатом анализа на этапе (140) является набор выводов, формируемых в зависимости от результата прохождения каждого правила. Под выводом понимается позитивное (формируемое в случае успешного прохождения правил) или негативное (формируемое в случае отрицательного результата прохождения правил) сообщение. В случае реализации негативного сценария сообщение содержит соответствующие пояснения о его причинах. Например, в отношении выписки из ЕГРЮЛ проводится проверка, в том числе, на наличие в ней атрибута с полным наименованием юридического лица, являющегося обязательным для данного документа. При отсутствии данного атрибута, выводится негативное сообщение с пояснением "В ЕГРЮЛ отсутствует полное наименование ЮЛ" (фиг. 6).

Наборы правил, сгруппированных в зависимости от специфики проводимой проверки, формируют блоки проверки. Примерами блоков могут служить блок по проверке клиента по выписке из ЕГРЮЛ или блок проверки срока действия общества. Пример представлен ниже:

Раздел 1 "Проверка правоспособности".

Блок 1 "Проверка правоспособности на основании данных ЕГРЮЛ".

Блок 2 "Проверка срока действия общества по Уставу".

На этапе (140) выполняется также валидация документов. Под валидацией понимается процесс первичного отбора документов для последующей проверки. Например, производится определение принадлежности документа Обществу и наличия в документе императивных атрибутов (атрибутов, наличие которых в документе обязательно в силу закона). Из сформированного списка валидированных документов исключаются документы, утратившие юридическую силу, и документы, юридическая проверка которых в рамках формирования данного заключения не требуется. Например, среди всех найденных протоколов о назначении единоличного исполнительного органа для проведения дальнейшей проверки будет оставлен только актуальный (протокол с датой, наиболее близкой к дате проверки).

Если на этапе (120) было получено несколько документов для обработки запроса (110), то может выполняться кросс-проверка их атрибутов. Под кросс-проверкой атрибутов понимается сравнение юридических фактов, содержащихся в различных документах, между собой. Например, кросс-проверка

должности единоличного исполнительного органа, указанной в уставе и в протоколе, производится путем проверки совпадения информации, содержащейся в атрибутах, которые были получены в результате извлечения наименования соответствующей должности из протокола и устава соответственно. Факт извлечения атрибута означает наличие соответствующего юридического факта в документе. Примером проверки факта извлечения атрибута может служить выявление атрибута, свидетельствующего о нахождении в залоге долей учредителей клиента. В данном случае само отсутствие либо наличие данного атрибута влияет на развитие дальнейшей логики проверки.

Сопоставление атрибута с каким-либо критерием осуществляется, как правило, но не ограничиваясь, с использованием логических операторов сравнения ("больше", "меньше") или контекстного поиска ("содержит"). Например, вхождение в состав атрибута, отвечающего за наименование должности единоличного исполнительного органа, фразы "Индивидуальный предприниматель" влияет на развитие дальнейшей логики проверки. В качестве примера задачи анализа извлеченных атрибутов можно рассмотреть начальный этап алгоритма поиска комплекта устава и изменений к нему, являющегося актуальным на момент проведения проверки. На данном этапе проверки отбираются уставы и изменения, принадлежащие анализируемому клиенту. Отбор производится путем проверки совпадения атрибутов, содержащих основной государственный регистрационный номер (ОГРН) и государственный регистрационный номер (ГРН), извлеченных со штампа регистрирующего органа (далее штамп) с аналогичными атрибутами, извлеченными из единого государственного реестра юридических лиц (ЕГРЮЛ). При этом учитываются случаи, когда ОГРН и ГРН на штампе перепутаны местами, один из номеров отсутствует или в обоих графах штампа указан один и тот же номер. Также при проверке учитываются возможные опечатки в указанных номерах.

На этапе (150) выполняется формирование правового решения на основании проведенной проверки набора атрибутов. На данном этапе система автоматического принятия решения формирует юридическое заключение по результатам проведенных проверок набора извлеченных сущностей и атрибутов. Данное заключение содержит в том числе список использованных атрибутов, набор сообщений, сформированных в рамках проведенной проверки и информацию о положительном или отрицательном результате прохождения проверки в рамках каждого блока правил. Под отрицательным результатом понимается выявление системой автоматического принятия решения следующих обстоятельств, но не ограничиваясь: 1) необходимость получения от клиента дополнительных документов (в этом случае клиенту направляется список требуемых документов); 2) необходимость привлечения к проведению данной проверки человека-эксперта; 3) необходимость исключения клиента из дальнейшей автоматизированной проверки в виду выявленных в отношении него ограничений. Примером таких ограничений может служить нахождение клиента в стадии банкротства или ликвидации. Во всех остальных случаях результат проверки блока правил считается положительным. В примере алгоритма поиска актуального комплекта устава и изменений, рассмотренном на этапе (140), сформированное правовое решение будет включать в себя перечень использованных атрибутов, содержащих ОГРН, ГРН и даты, указанные в уставах, изменениях и в выписке из ЕГРЮЛ (фиг. 7). Если актуальный комплект был найден (положительный результат), то в заключении будет содержаться соответствующее сообщение с указанием дат регистрации актуального устава и изменений. Если же, по каким-то причинам, актуальный комплект найден не был (отрицательный результат), то заключение будет включать в себя запрос актуального комплекта, состоящего из устава и изменений (например, если найденный комплект не совпал с соответствующим списком изменений в выписке из ЕГРЮЛ), или вывод о необходимости проведения определенного этапа проверки с участием человека-эксперта (например, если во всех уставах в комплекте атрибуты, содержащие ОГРН, не совпадают с атрибутом, содержащим ОГРН, в выписке из ЕГРЮЛ).

На фиг. 8 представлен пример организации системы (200) для применения вышеописанного способа принятия юридических решений (100). Исполняемая программная логика способа (100) может располагаться на централизованном вычислительном устройстве, например, сервере (220), доступ к которому обеспечивается посредством сети передачи данных, в частности, сети "Интернет", с помощью устройства пользователя (210).

Запрос на получение той или иной услуги, требующей предоставления одного или нескольких документов (20) пользователем (10), осуществляется через соответствующий ресурс (веб-портал, мобильное приложение и т.п.), взаимодействие с которым осуществляется с помощью устройства пользователя (210). В качестве устройства (210) может применяться широкий спектр вычислительных устройств, например, персональный компьютер, ноутбук, смартфон, планшет, смарт-ТВ и т.п. При выборе требуемого типа услуги клиентом (10), приложение на ресурсе с помощью вычислительного устройства (210) в ответ на запрос клиента (10) направляет информацию на устройство (210) с перечнем необходимых документов (20), которые будут требоваться для дальнейшего оказания услуги, и проверка которых будет осуществляться с помощью исполнения способа (100) на сервере (220).

Передача документов (20) с помощью устройства (210) может осуществляться с помощью стандартной формы и принципов передачи данных, например, e-mail, мессенджеры, фотографическая съемка камерой устройства (210) с помощью специализированной функции мобильного или веб приложения (например, приложение Сбербанк Онлайн) и т.п. Документы (20), как правило, предоставляются в

виде цифровых изображений (скан копий, фотоснимков) в одном из поддерживаемых форматов, например, JPG, JPEG, TIFF, PNG и т.п.

На фиг. 9 представлена блок-схема процесса обработки документов (300) на сервере (220) с помощью выполнения алгоритмической работы способа (100). Обработка документов осуществляется при их поступлении на сервер (220) в процессе требуемого анализа по формированию юридического заключения.

Обработка полученных документов (302) клиента (10) выполняется на основании поступающего запроса (301) с помощью работы способа (100), описанного ранее. По итогам проверки на этапе (303) осуществляется формирование юридического заключения, которое далее передается на устройство (210) клиента (10) или в другие смежные системы. Формируемое заключение (303) может содержать информацию о положительном или отрицательном итоге проверки документов. В частности, заключение может содержать сведения о соответствии документов заданным требованиям (например, для оказания услуги), либо, содержать информацию о необходимости совершения дополнительных действий.

При отрицательном результате проверки дополнительными действиями могут служить, например, предоставление дополнительных документов или предоставление ранее представленных документов в более высоком качестве. Также, на этапе (303) может формироваться отрицательное решение по представленным документам (20) на основании их несоответствия нормам, установленным законом и/или ВНД.

Формирование юридического заключения (303) может выполняться для широкого профиля задач, например, получение кредита, открытие вклада, выпуск карт оплаты, заключение сделок (например, с помощью смарт-контрактов и т.п.), консультационная проверка договора/контракта и т.п.

На фиг. 10 представлен пример общего вида вычислительного устройства (400), на базе которого может быть реализована одна или несколько автоматизированных систем, обеспечивающих реализацию заявленного способа (100) или (300). Упомянутые в настоящих материалах заявки такие устройства, как сервер, устройство пользователя и т.д. могут выполняться полностью или частично на базе устройства (400).

В общем случае, устройство (400) содержит объединенные общей шиной информационного обмена (410) один или несколько процессоров (401), средства памяти, такие как ОЗУ (402) и ПЗУ (403), интерфейсы ввода/вывода (404), устройства ввода/вывода (405), и устройство для сетевого взаимодействия (406).

Процессор (401) (или несколько процессоров, многоядерный процессор и т.п.) может выбираться из ассортимента устройств, широко применяемых в настоящее время, например, таких производителей, как: Intel™, AMD™, Apple™, Samsung Exynos™, MediaTEK™, Qualcomm Snapdragon™ и т.п. Под процессором или одним из используемых процессоров в устройстве (400) также необходимо учитывать графический процессор, например, GPU NVIDIA или Graphcore, тип которых также является пригодным для полного или частичного выполнения способа (100), а также может применяться для обучения и применения моделей машинного обучения в различных информационных системах.

ОЗУ (402) представляет собой оперативную память и предназначено для хранения исполняемых процессором (401) машиночитаемых инструкций для выполнения необходимых операций по логической обработке данных. ОЗУ (402), как правило, содержит исполняемые инструкции операционной системы и соответствующих программных компонент (приложения, программные модули и т.п.). При этом, в качестве ОЗУ (402) может выступать доступный объем памяти графической карты или графического процессора.

ПЗУ (403) представляет собой одно или более устройств постоянного хранения данных, например, жесткий диск (HDD), твердотельный накопитель данных (SSD), флэш-память (EEPROM, NAND и т.п.), оптические носители информации (CD-R/RW, DVD-R/RW, BlueRay Disc, MD) и др.

Для организации работы компонентов устройства (400) и организации работы внешних подключаемых устройств применяются различные виды интерфейсов В/В (404). Выбор соответствующих интерфейсов зависит от конкретного исполнения вычислительного устройства, которые могут представлять собой, не ограничиваясь: PCI, AGP, PS/2, IrDa, FireWire, LPT, COM, SATA, IDE, Lightning, USB (2.0, 3.0, 3.1, micro, mini, type C), TRS/Audio jack (2.5, 3.5, 6.35), HDMI, DVI, VGA, Display Port, RJ45, RS232 и т.п. Для обеспечения взаимодействия пользователя с устройством (400) применяются различные средства (405) В/В информации, например, клавиатура, дисплей (монитор), сенсорный дисплей, тач-пад, джойстик, манипулятор, мышь, световое перо, стилус, сенсорная панель, трекбол, динамики, микрофон, средства дополненной реальности, оптические сенсоры, планшет, световые индикаторы, проектор, камера, средства биометрической идентификации (сканер сетчатки глаза, сканер отпечатков пальцев, модуль распознавания голоса) и т.п.

Средство сетевого взаимодействия (406) обеспечивает передачу данных посредством внутренней или внешней вычислительной сети, например, Интранет, Интернет, ЛВС и т.п. В качестве одного или более средств (406) может использоваться, но не ограничиваясь: Ethernet карта, GSM модем, GPRS модем, LTE модем, 5G модем, модуль спутниковой связи, NFC модуль, Bluetooth и/или BLE модуль, Wi-Fi модуль и др. Представленные материалы заявки раскрывают предпочтительные примеры реализации

технического решения и не должны трактоваться как ограничивающие иные, частные примеры его воплощения, не выходящие за пределы испрашиваемой правовой охраны, которые являются очевидными для специалистов соответствующей области техники.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Компьютерно-реализуемый способ автоматического принятия правового решения, выполняемый с помощью по меньшей мере одного процессора, содержащий этапы, на которых получают запрос на формирование правового решения;
получают по меньшей мере один документ, соответствующий упомянутому запросу;
обрабатывают полученный документ на предмет выявления в документе по меньшей мере одной сущности, относящейся к запросу правового решения, и осуществляют ее извлечение из документа с помощью алгоритма машинного обучения;
осуществляют юридический анализ упомянутой по меньшей мере одной извлеченной сущности с помощью набора правил, сформированных с помощью предметно-ориентированного языка;
формируют правовое решение на основании проведенного юридического анализа упомянутых извлеченных сущностей упомянутого по меньшей мере одного документа.
2. Способ по п.1, характеризующийся тем, что на этапе обработки документа осуществляется распознавание печатного текста документа, при котором выполняется автоматическое распознавание изображений документов и извлечение текстовых данных из них.
3. Способ по п.2, характеризующийся тем, что на основании извлеченных текстовых данных формируется новый документ.
4. Способ по п.3, характеризующийся тем, что выполняется определение типа документа с помощью сравнения извлеченных текстовых данных.
5. Способ по п.1, характеризующийся тем, что при получении нескольких документов осуществляется их обработка на предмет объединения в рамках полученного запроса формирования правового решения.
6. Способ по п.3, характеризующийся тем, что выполняется сравнение изображения печатного текста документа с новым документом, сформированным по итогам извлечения текстовых данных, сохраненным в новом документе, образованном в результате распознавания.
7. Способ по п.6, характеризующийся тем, что при наличии расхождения между документами в результате сравнения, в новом документе выполняют исправления ошибок распознавания изображения документа.
8. Способ по п.1, характеризующийся тем, что дополнительно выполняют верификацию по меньшей мере одной извлеченной сущности из документа.
9. Способ по п.8, характеризующийся тем, что выполняется анализ смыслового соответствия по меньшей мере одной извлеченной сущности участку текста документа, из которого было осуществлено упомянутое извлечение.
10. Способ по п.8, характеризующийся тем, что выполняется корректировка по меньшей мере одной извлеченной сущности, причем такая корректировка включает в себя по меньшей мере одно из: исправление орфографических ошибок, пунктуации, технических ошибок, смысловых ошибок.
11. Способ по п.5, характеризующийся тем, что выполняется фильтрация документов, по меньшей мере, утративших юридическую силу.
12. Способ по п.5, характеризующийся тем, что на этапе выполняется кросс-проверка юридических сущностей, при которой осуществляется сравнение юридических фактов, содержащихся в документах, отобранных для формирования юридического заключения.
13. Устройство автоматического принятия правового решения, содержащее по меньшей мере один процессор и по меньшей мере одно средство памяти, соединенное с процессором и содержащее машиночитаемые инструкции, которые при их исполнении процессором обеспечивают выполнение вышеуказанного способа по любому из пп.1-12.
14. Система автоматического принятия правового решения, содержащая систему интеллектуального распознавания документов, включающую
модуль обработки, выполняющий
получение запроса на формирование правового решения,
получение по меньшей мере одного документа, связанного с упомянутым запросом;
модуль извлечения сущности, обеспечивающий выявление в документе по меньшей мере одной сущности, относящейся к запросу правового решения, и ее извлечение из документа с помощью алгоритма машинного обучения;
систему автоматизированного принятия решений, обеспечивающую выполнение юридического анализа упомянутой по меньшей мере одной извлеченной сущности с помощью набора правил, сформированных с помощью предметно-ориентированного языка;
формирование правового решения на основании проведенного юридического анализа упомянутых

извлеченных сущностей упомянутого по меньшей мере одного документа.

15. Система по п.14, характеризующаяся тем, что дополнительно содержит модуль верификации извлеченных сущностей, выполняющий проверку правильности извлечения атрибутов документов.

16. Система по п.15, характеризующаяся тем, что модуль верификации выполняет анализ смыслового соответствия по меньшей мере одной извлеченной сущности участку текста документа, из которого было осуществлено упомянутое извлечение.

17. Система по п.15, характеризующаяся тем, что модуль верификации выполняет корректировку по меньшей мере одной извлеченной сущности, причем такая корректировка включает в себя по меньшей мере одно из: исправление орфографических ошибок, пунктуации, технических ошибок, смысловых ошибок.

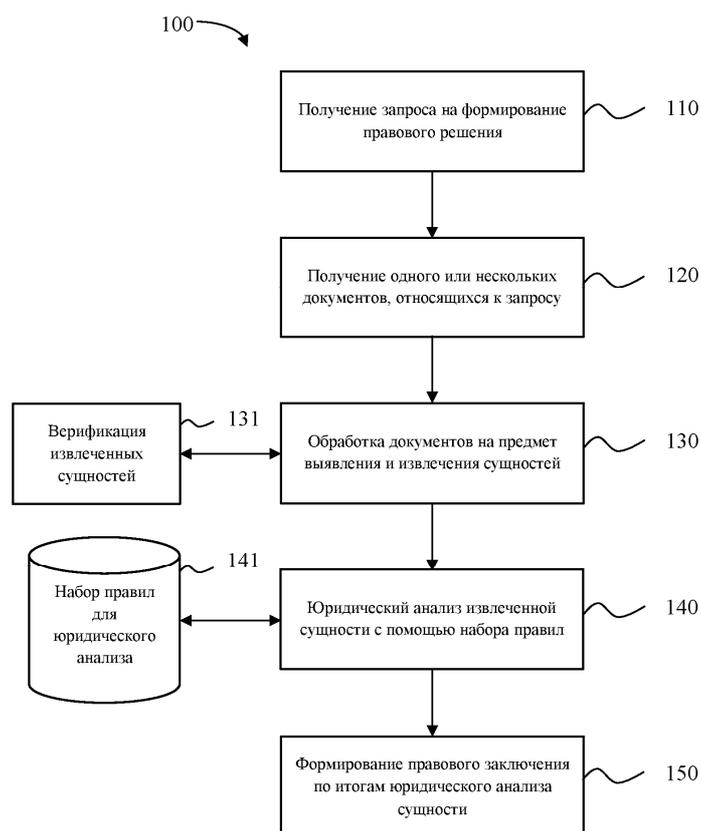
18. Система по п.14, характеризующаяся тем, что модуль обработки выполняет распознавание печатного текста документа, при котором выполняется автоматическое распознавание изображения документов и извлечение текстовых данных из них.

19. Система по п.18, характеризующаяся тем, что на основании извлеченных текстовых данных формируется новый документ.

20. Система по п.18, характеризующаяся тем, что выполняется определение типа документа с помощью сравнения извлеченных текстовых данных.

21. Система по п.19, характеризующаяся тем, что выполняется сравнение изображения печатного текста документа с новым документом, сформированным по итогам извлечения текстовых данных, сохраненным в новом документе, образованном в результате распознавания.

22. Система по п.21, характеризующаяся тем, что при наличии расхождения между документами в результате сравнения, в новом документе выполняют исправления ошибок распознавания изображения документа.



Фиг. 1

1.6. Местонахождение Общества определяется местом его государственной регистрации: Российская Федерация, 152020, Ярославская область, город Переславль-Залесский, улица

Фиг. 2

```

- <Field>
  <Name>СубъектРФ</Name>
  <FullName>Устав с изменениями\МН\СубъектРФ[0]</FullName>
  <Text>Ярославская</Text>
</Field>
- <Field>
  <Name>НасПункт</Name>
  <FullName>Устав с изменениями\МН\НасПункт</FullName>
  <Text>Переславль-Залесский</Text>
</Field>

```

Фиг. 3

Разметка сущности в тексте:

```

? 3 Место нахождения Общества - 672000 Российская Федерация [1 5_МНОское Забайкальский край г Чита] ул Чкалова 158
? 4 Почтовый адрес Общества- 672050 Российская Федерация Забайкальский край г Чита Главпочтамт а я 371

```

Визуализация результата разметки:

Забайкальский край г Чита

Группы сущностей 1 5_МНОское

Все

0 Мигрирующие сущности 1: СубъектРФ Забайкальский край

1 Общие положения

10 Крупная сделка 2: НасПункт

11 Сделка с заинтересованностью Чита

Фиг. 4

Дочерние документы

Наименование документа	Создатель	Дата создания
Копии протоколов (решений) об избрании органов управления_ИРД.pdf	ectm_support_4_ird	15.11.2019 11:47:32
Копии протоколов (решений) об избрании органов управления_ИРД.xml	ectm_support_4_ird	15.11.2019 11:47:32

Фиг. 5

Если не существует Документы_ЕГРЮЛ.СвЮЛ.СвНаимЮЛ.НаимЮЛПолн то
 Вывод Выписка из ЕГРЮЛ не валидирована Требуется юридическая экспертиза "В ЕГРЮЛ отсутствует полное наименование ЮЛ"

Фиг. 6

Общие вывод и сведения о контрагенте

Идентификатор правового запроса АС ПО

IA-57877764

Дата заключения

06.11.2019

Решение

Положительное заключение, формирование КОД возможно

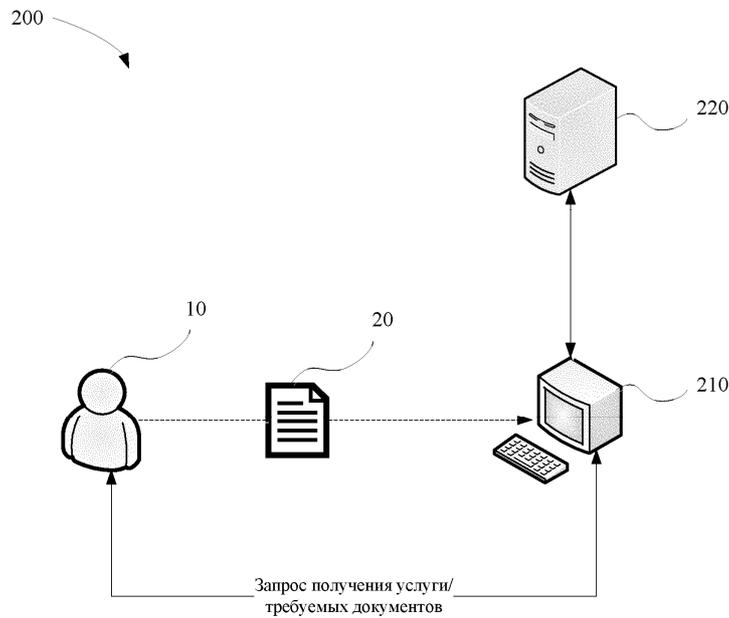
Проверяемый контрагент

Роль контрагента	Ю СЯМ	ИНН	КПП	ОГРН	Полное наименование	Сокращённое наименование
Заемщик	I-145YR	7716801406	771026763	1093050695506	ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ "РОМАШКА "	ООО "РОМАШКА"

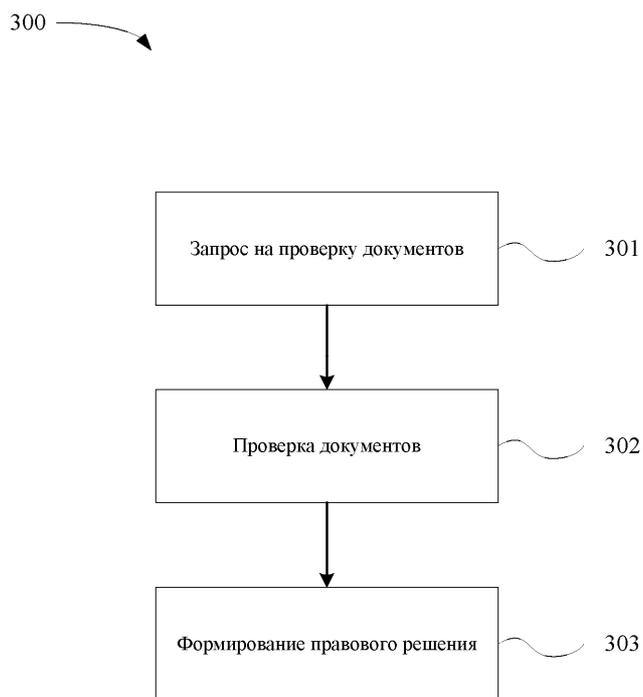
1. Выводы по контрагенту

1	Правоспособность подтверждена	Да
2	Полномочия подтверждены	Да
3	Полное ФИО лица, заключающего договор	СЕМИЦВЕТОВ ЕВГЕНИЙ АЛЕКСЕЕВИЧ
3.1.	Должность в соответствии с правоустанавливающими документами	ГЕНЕРАЛЬНЫЙ ДИРЕКТОР
3.2.	ИНН ЕИО	772891528198
3.3.	Срок истечения полномочий	06.09.2020
4	Необходимость запроса корпоративного одобрения	
4.1.	Заинтересованность	Нет
4.1.1.	Одобряющий орган	
4.2.	Корпоративные ограничения	Нет
5	Правовой риск	Уровень риска
	<i>Правовой риск отсутствует</i>	

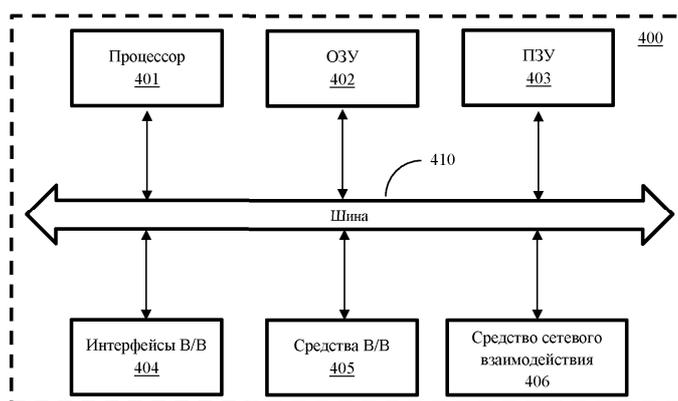
Фиг. 7



Фиг. 8



Фиг. 9



Фиг. 10