



ФЕДЕРАЛЬНАЯ СЛУЖБА  
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

## (12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК

*G06K 9/627* (2019.02); *G06N 7/005* (2019.02)

(21)(22) Заявка: 2018110380, 23.03.2018

(24) Дата начала отсчета срока действия патента:  
23.03.2018Дата регистрации:  
23.07.2019

Приоритет(ы):

(22) Дата подачи заявки: 23.03.2018

(45) Опубликовано: 23.07.2019 Бюл. № 21

Адрес для переписки:

127273, Москва, а/я 20, ООО "Аби Продакшн",  
для Марья С.В.

(72) Автор(ы):

Каленков Максим Петрович (RU)

(73) Патентообладатель(и):

Общество с ограниченной ответственностью  
"Аби Продакшн" (RU)

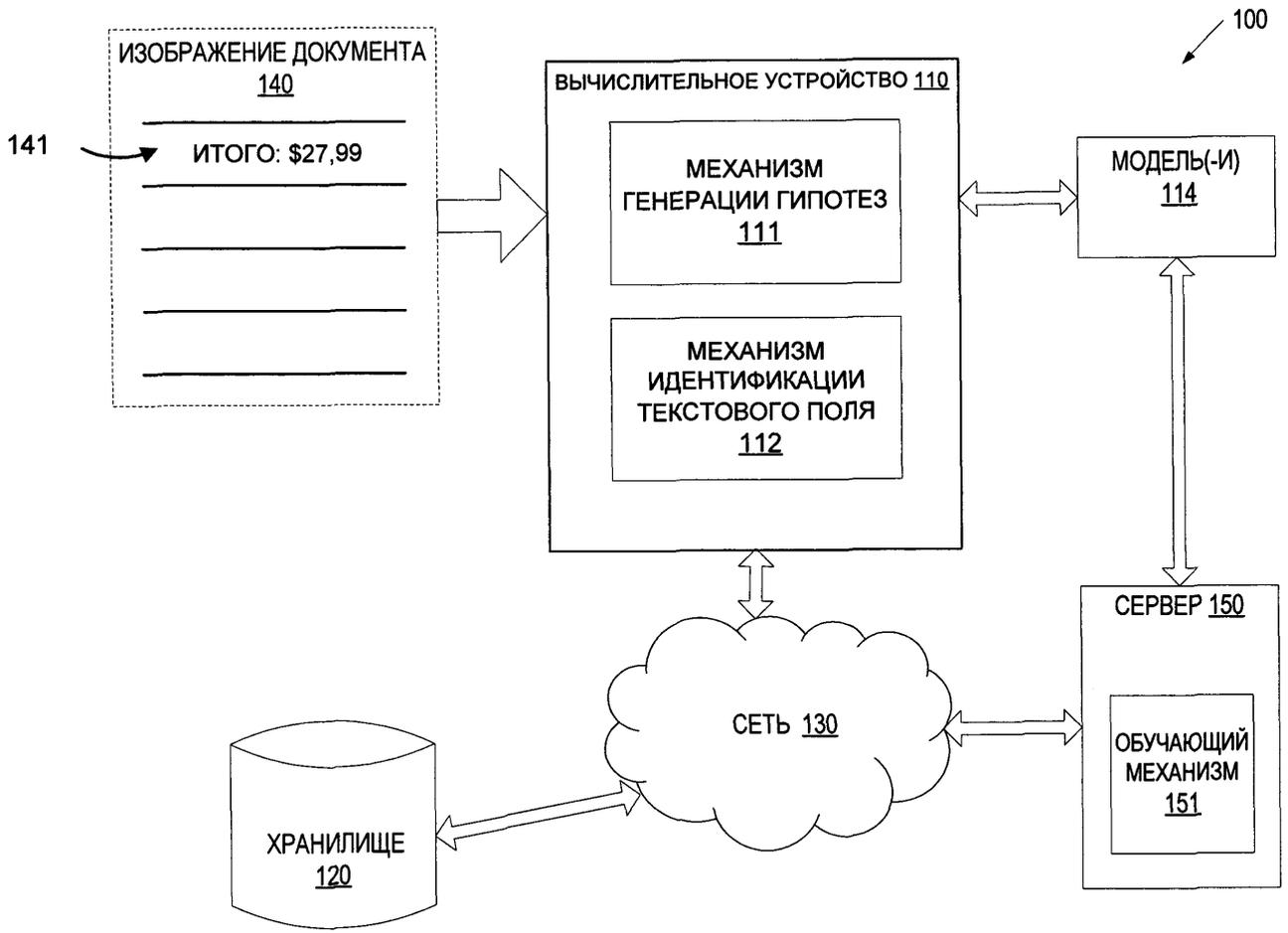
(56) Список документов, цитированных в отчете

о поиске: RU 2613846 C2, 21.03.2017. RU  
2635259 C1, 09.11.2017. US 2014/201126 A1,  
17.07.2014. US 2009/116755 A1, 07.05.2009. US  
2010/017 A1, 21.01.2010.(54) ИДЕНТИФИКАЦИЯ ПОЛЕЙ НА ИЗОБРАЖЕНИИ С ИСПОЛЬЗОВАНИЕМ  
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

(57) Реферат:

Изобретение относится к механизму идентификации текстового поля. Технический результат заключается в расширении арсенала средств для идентификации текстовых полей. В способе получают одну или более гипотез для типа поля первого текстового поля, присутствующего на изображении документа, и создают трехмерную матрицу признаков, представляющую часть изображения,

содержащую первое поле. Механизм идентификации текстового поля предоставляет трехмерную матрицу признаков в качестве входных данных для обученной модели машинного обучения и получает выходные данные из обученной модели машинного обучения, при этом выходные данные содержат оценку качества одной или более гипотез. 3 н. и 17 з.п. ф-лы, 8 ил.



Фиг. 1

RU 2695489 C1

RU 2695489 C1



FEDERAL SERVICE  
FOR INTELLECTUAL PROPERTY

(12) **ABSTRACT OF INVENTION**

(52) CPC  
*G06K 9/627* (2019.02); *G06N 7/005* (2019.02)

(21)(22) Application: **2018110380, 23.03.2018**

(24) Effective date for property rights:  
**23.03.2018**

Registration date:  
**23.07.2019**

Priority:

(22) Date of filing: **23.03.2018**

(45) Date of publication: **23.07.2019** Bull. № 21

Mail address:  
**127273, Moskva, a/ya 20, OOO "Abi Prodakshn",  
dlya Mareya S.V.**

(72) Inventor(s):

**Kalenkov Maksim Petrovich (RU)**

(73) Proprietor(s):

**Obshchestvo s ogranichennoj otvetstvennostyu  
"Abi Prodakshn" (RU)**

(54) **IDENTIFICATION OF FIELDS ON AN IMAGE USING ARTIFICIAL INTELLIGENCE**

(57) Abstract:

FIELD: physics.

SUBSTANCE: invention relates to a text field identification mechanism. Method includes obtaining one or more hypotheses for the type of the field of the first text field present on the image of the document and creating a three-dimensional matrix of features which represents part of the image containing the first field. Text field identification mechanism provides a

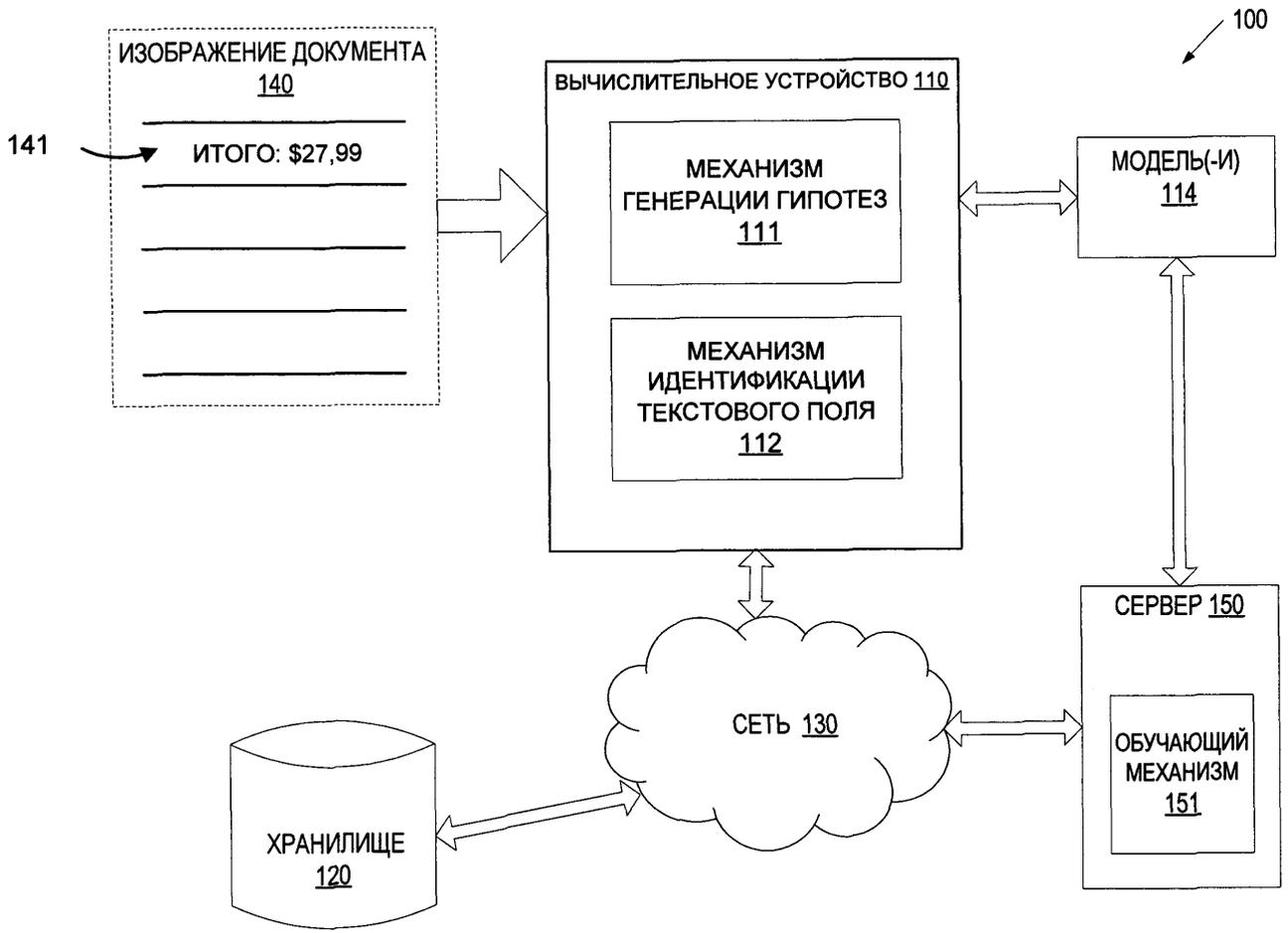
three-dimensional feature matrix as input data for the trained machine learning model and obtains output data from the trained machine learning model, wherein the output data contain a quality estimate of one or more hypotheses.

EFFECT: technical result consists in expansion of arsenal of means for identification of text fields.

20 cl, 8 dwg

**RU 2 695 489 C 1**

**RU 2 695 489 C 1**



Фиг. 1

RU 2695489 C1

RU 2695489 C1

## ОБЛАСТЬ ТЕХНИКИ

[001] Настоящее изобретение в целом относится к вычислительным системам, а в частности - к системам и способам идентификации текстовых полей на основе контекста с использованием искусственного интеллекта, включая сверточные нейронные сети.

## УРОВЕНЬ ТЕХНИКИ

[002] Извлечение информации может включать в себя анализ текста на естественном языке для распознавания и классификации информационных объектов в соответствии с заранее определенным набором категорий (таких как имена лиц, организации, местоположения, выражения времени, количества, денежные значения, проценты, и т.д.). Извлечение информации может дополнительно идентифицировать отношения между распознанными именованными объектами и/или другими информационными объектами.

## КРАТКОЕ ИЗЛОЖЕНИЕ СУЩНОСТИ ИЗОБРЕТЕНИЯ

[003] В одном из вариантов реализации изобретения механизм идентификации текстового поля получает одну или более гипотез для типа поля первого текстового поля, присутствующего на изображении документа. В одном из вариантов реализации изобретения механизм идентификации текстового поля обрабатывает изображение для создания трехмерной матрицы признаков, представляющей часть изображения, содержащую первое поле. Для этого механизм идентификации текстового поля может определять множество горизонтальных строк текста, присутствующих на изображении, в котором одна из множества горизонтальных строк содержит первое поле, задавать систему координат для множества горизонтальных строк и сдвигать систему координат по горизонтали на основании положения первого поля на изображении для формирования смещенной системы координат, в которой трехмерная матрица признаков основана на смещенной системе координат. Для задания системы координат, механизм идентификации текстового поля может находить на изображении левый и правый края документа, связывать первое значение с первым положением на пересечении левого края и, по меньшей мере, с одной из множества горизонтальных строк, а также связывать второе значение со вторым положением на пересечении правого края и, по меньшей мере, с одной из множества горизонтальных строк. Чтобы сдвинуть систему координат по горизонтали, механизм идентификации текстового поля может сдвинуть первое значение в положение первого поля изображения.

[004] В одном из вариантов реализации изобретения, механизм идентификации текстового поля дополнительно кадрирует изображение для формирования кадрированного изображения, содержащего заданное количество строк выше и ниже одной из множества горизонтальных строк, которая содержит первое поле, разбивает кадрированное изображение на множество ячеек и вычисляет множество признаков для каждого из множества ячеек, в котором множество признаков содержит информацию, относящуюся к графическим элементам, представляющим один или более символов, присутствующих в соответствующей ячейке, и содержит, по меньшей мере, один компонент трехмерной матрицы признаков.

[005] В одном из вариантов реализации изобретения, механизм идентификации текстового поля предоставляет трехмерную матрицу признаков в качестве входных данных для обученной модели машинного обучения и получает выходные данные из обученной модели машинного обучения. Обученная модель машинного обучения может содержать, например, сверточную нейронную сеть. Выходные данные из обученной модели машинного обучения содержат оценку качества одной или более гипотез. Эта оценка содержит, по меньшей мере, одно из: указание, что первая гипотеза из одной

или более гипотез является предпочтительной гипотезой из множества гипотез, или значение уверенности, связанное с одной или более гипотезами. В одном из вариантов реализации изобретения обученная модель машинного обучения обучается с использованием обучающей выборки данных, содержащей примеры изображений документов, содержащих одно или более полей в качестве вводных данных для обучения, и один или более идентификаторов типа поля, который правильно соответствует одному или более полям в качестве целевых выводных данных.

#### КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

[006] Для более полного понимания настоящего изобретения ниже приводится подробное описание, в котором для примера, а не способом ограничения, оно иллюстрируется со ссылкой на чертежи, на которых:

[007] На Фиг. 1 изображена схема компонентов верхнего уровня для примера архитектуры системы в соответствии с одним или более вариантами реализации настоящего изобретения.

[008] На Фиг. 2А и 2В приведено изображение документа, имеющее количество полей, подлежащих идентификации в соответствии с одним или более вариантами реализации настоящего изобретения.

[009] На Фиг. 3 приведена блок-схема, иллюстрирующая способ идентификации поля в соответствии с одним или более вариантами реализации настоящего изобретения.

[0010] На Фиг. 4 показана блок-схема, иллюстрирующая способ обработки изображения документа в соответствии с одним или более вариантами реализации настоящего изобретения.

[0011] На Фиг. 5 показан пример системы координат для горизонтальных текстовых строк на изображении документа в соответствии с одним или более вариантами реализации настоящего изобретения.

[0012] На Фиг. 6 приведены геометрические признаки множества полей на изображении документа в соответствии с одним или более вариантами реализации настоящего изобретения.

[0013] На Фиг. 7 показана топология сети для оценки уверенности гипотезы типа поля на изображении документа в соответствии с одним или более вариантами реализации настоящего изобретения.

[0014] На Фиг. 8 приведен пример вычислительной системы, которая может выполнять один или более способов, описанных в настоящем документе, в соответствии с одним или более вариантами реализации настоящего изобретения.

#### ПОДРОБНОЕ ОПИСАНИЕ

[0015] Описаны варианты идентификации текстовых полей на основе контекста с использованием искусственного интеллекта, включая сверточные нейронные сети. Одним из алгоритмов идентификации полей на изображении документа является эвристический подход. В эвристическом подходе рассматривается большое количество (порядка сотен) изображений документов, таких, например, как ресторанные чеки или счета, и накапливается статистика относительно того, какой текст (например, ключевые слова) используется рядом с определенным полем и где этот текст может быть расположен относительно поля (например, справа, слева, выше, ниже). Например, эвристический подход отслеживает, какое слово или слова обычно расположены рядом с полем, указывающим общую сумму покупки, какое слово или слова находятся рядом с полем, указывающим на применимые налоги, какое слово или слова написаны рядом с полем, указывающим общую сумму оплаты по кредитной карте и т.д. На основе этих статистических данных при обработке изображения нового чека можно определить,

какие данные, обнаруженные на изображении документа, соответствуют определенному полю. Эвристический подход не всегда работает точно, потому что, если по какой-то причине чек был распознан с ошибками, а именно в словосочетаниях «ОБЩИЙ НАЛОГ» и «ОБЩИЙ ПЛАТЕЖ» слова «налог» и «платеж» были распознаны плохо, то соответствующие значения могут быть неправильно классифицированы.

[0016] Другим подходом для идентификации полей является метод распознавания именованных сущностей (NER, Named Entity Recognition). В этом подходе, после получения всего распознанного текста изображения документа, текст разбивается на отдельные слова, которые подаются на вход рекуррентной нейронной сети. Сеть определяет вероятность того, что каждое слово соответствует определенному классу, который в случае чеков является конкретным полем. Качество определения метода NER обычно измеряется на основе найденных и пропущенных слов или символов. Но при поиске полей в чеке представляют интерес соответствующие значения полей. То есть после того как выделен текст поля, также необходимо извлечь значение поля. В целом метод NER работает хорошо, хотя и не так хорошо, как некоторые известные специализированные методы, которые извлекают определенные поля, используя все данные, специфичные для этих полей, включая геометрию, контекст и арифметические правила.

[0017] В одном из вариантов реализации изобретения описанные здесь методы идентификации поля содержат создание одной или нескольких гипотез относительно типа поля для конкретного поля на изображении документа (например, чека). Для исходных гипотез можно использовать простой механизм поиска полей с помощью регулярных выражений. Поиск регулярного выражения может использоваться, чтобы отличить разные типы данных в чеке, например, чтобы отличить денежные суммы от телефонных номеров, но не поможет различить другие типы более похожих данных (например, различные типы денежных сумм, такие как итого, сдача, оплата по банковской карте, применяемая скидка и т.д.). В дополнение к регулярным выражениям шаблоны могут использоваться для идентификации разных полей в чеке. Шаблоны могут хранить информацию о структуре чека конкретного поставщика, включая ожидаемый тип поля, связанный с расположением поля в чеке. Однако, одно поле или целые строки шаблона могут плохо накладываться на конкретный чек из-за ошибок распознавания или локальных отличий конкретного чека от чеков, используемых при обучении шаблона. Таким образом, в обоих случаях следующим шагом после принятия одной или нескольких гипотез является оценка качества гипотез для индивидуальных полей.

[0018] Описанное здесь представляет собой систему и способ оценки гипотез для конкретных полей. Если существует несколько гипотез то, в зависимости от варианта реализации изобретения, способ может выбрать наилучшую гипотезу (то есть, наиболее правильную) или отсортировать множественные гипотезы путем оценки качества. Если существует только одна гипотеза, способ может оценить значение уверенности гипотезы, чтобы указать, насколько вероятно выбранная для поля гипотеза верна. В результате такой оценки способ может предоставить клиенту не только результаты поиска поля, но и указание уверенности результата.

[0019] Варианты реализации настоящего изобретения проводят такую оценку путем использования набора моделей машинного обучения (например, нейронных сетей) для эффективной идентификации текстовых полей на изображении. Набор моделей машинного обучения может обучаться на группе изображений документов, которые формируют обучающую выборку данных. Обучающая выборка данных содержит

примеры изображений документов, включающих в себя одно или более полей в качестве вводных данных для обучения, и один или более идентификаторов типа поля, который правильно соответствует одному или более полям в качестве целевых выходных данных.

5 [0020] В настоящем документе могут попеременно использоваться термины «символ», «буква» и «кластер». Кластер может означать элементарный неделимый графический элемент (например, графемы или лигатуры), который связывается общим логическим значением. Кроме того, термин «слово» может означать последовательность символов, а термин «предложение» может означать последовательность слов.

10 [0021] После обучения набор моделей машинного обучения может использоваться для идентификации текстовых полей и для выбора типа поля с наибольшей уверенностью для конкретного поля. Использование моделей машинного обучения (например, сверточных нейронных сетей) избавляет от необходимости ручной разметки ключевых слов для поиска полей в чеке, поскольку ручная работа заменяется машинным обучением. Описанные здесь методы позволяют использовать простую топологию  
15 сети, и сеть быстро обучается на относительно небольшом наборе данных, например, по сравнению с NER. Дополнительно этот метод легко применяется для нескольких случаев использования, и сеть может быть обучена с использованием чеков одного поставщика, а затем применяться к чекам другого поставщика с высоким качеством результатов. Более того, использование сверточной сети позволяет уменьшить  
20 количество ошибок при поиске полей на изображении чеков примерно на 5-30%.

[0022] На Фиг. 1 изображена диаграмма компонентов верхнего уровня для пояснения архитектуры системы 100 в соответствии с одним или более вариантами реализации  
настоящего изобретения. Архитектура системы 100 содержит вычислительное устройство  
25 110, хранилище 120 и сервер 150, подключенный к сети 130. Сеть 130 может быть общественной сетью (например, Интернет), частной сетью (например, локальная сеть (LAN, local area network) или распределенной сетью (WAN, wide area network)), а также их комбинацией.

[0023] Вычислительное устройство 110 может выполнять идентификацию поля с использованием искусственного интеллекта для эффективной идентификации и  
30 классификации одного или нескольких полей на изображении документа 140. Идентифицированные поля могут быть идентифицированы по одному или более словам и могут содержать одно или более значений. Каждое из идентифицированных слов может состоять из одного или более символов (например, кластеров). Вычислительное устройство 110 может быть настольным компьютером, портативным компьютером,  
35 смартфоном, планшетным компьютером, сервером, сканером или любым подходящим вычислительным устройством, способным использовать технологии, описанные в этом изобретении. Изображение документа 140, содержащее одно или более полей 141, может передаваться в вычислительное устройство 110. Следует отметить, что изображение документа 140 может содержать напечатанный или рукописный текст на любом языке.

40 [0024] Документ 140 может быть получен любым подходящим способом. Например, вычислительное устройство 110 может получить цифровую копию документа 140 путем сканирования документа или фотографирования документа. Кроме того, в тех вариантах реализации изобретения, где вычислительное устройство 110 представляет собой сервер, клиентское устройство, подключенное по сети 130 к серверу, может загружать цифровую  
45 копию документа 140 на сервер. В тех вариантах реализации изобретения, где вычислительное устройство 110 является клиентским устройством, соединенным с сервером по сети 130, клиентское устройство может загружать изображение документа 140 с сервера.

[0025] Изображение документа 140 может быть использовано для обучения множества моделей машинного обучения или может быть новым документом, для которого желательно выполнить идентификацию поля. Соответственно, на предварительных этапах обработки, изображения документа 140 можно подготовить для обучения набора 5 моделей машинного обучения или для последующей идентификации. Например, на изображении документа 140 может быть выбрано вручную или автоматически поле 141, могут быть отмечены символы, могут быть выпрямлены, масштабированы и (или) бинаризованы строки текста. Распрямление строк может быть выполнено до обучения набора моделей машинного обучения и (или) идентификации поля 141 на изображении 10 документа 140 для приведения строки текста к одинаковой высоте (например, 80 пикселей).

[0026] В одном из вариантов реализации изобретения вычислительное устройство 110 может содержать механизм генерации гипотез 111 и механизм идентификации текстового поля 112. Каждый из механизмов генерации гипотез 111 и идентификации 15 текстового поля 112 может содержать инструкции, сохраненные на одном или более физических машиночитаемых носителях данных вычислительного устройства 110 и выполняемые на одном или более устройствах обработки вычислительного устройства 110. В одном из вариантов реализации изобретения механизм генерации гипотез 111 выдвигает одну или более исходных гипотез, определяющих тип поля для поля 141. 20 Например, исходные гипотезы могут быть порождены используя простой механизм поиска полей с помощью регулярных выражений, используя шаблоны для определения разных полей в чеке. В одном из вариантов реализации изобретения механизм идентификации текстового поля 112 может использовать множество обученных моделей машинного обучения 114, которые обучены и используются для идентификации полей 25 на изображении документа 140 и подтверждают или опровергают исходные гипотезы. Механизм идентификации текстового поля 112 также может предварительно обрабатывать полученные изображения, такие как изображение документа 140, перед использованием этих изображений для обучения моделей машинного обучения 114 и (или) применения набора обученных моделей машинного обучения 114 к изображениям. 30 В некоторых вариантах реализации набор обученных моделей машинного обучения 114 может быть частью механизма идентификации текстового поля 112 или может быть доступен на другой машине (например, на сервере 150) через механизм идентификации текстового поля 112. Основываясь на выходных данных набора обученных моделей машинного обучения 114, механизм идентификации текстового поля 112 может получить 35 оценку качества одной или более гипотез для типа поля для поля 141 на изображении документа 140.

[0027] Сервером 150 может быть стоечный сервер, маршрутизатор, персональный компьютер, карманный персональный компьютер, мобильный телефон, портативный компьютер, планшетный компьютер, фотокамера, видеокамера, нетбук, настольный 40 компьютер, медицентр или их сочетание. Сервер 150 может содержать механизм обучения 151. Набор моделей машинного обучения 114 может ссылаться на артефакты моделей, созданные обучающим механизмом 151 с использованием обучающих данных, которые содержат обучающие входные данные и соответствующие целевые выходные данные (правильные ответы на соответствующие обучающие входные данные). В 45 процессе обучения могут быть найдены конфигурации в обучающих данных, которые преобразуют входные данные обучения в целевые выходные данные (ответ, который следует предсказать), и впоследствии могут быть использованы моделями машинного обучения 114 для будущих прогнозов. Как более подробно будет описано ниже, набор

моделей машинного обучения 114 может быть составлен, например, из одного уровня линейных или нелинейных операций (например, машина опорных векторов [SVM, support vector machine]) или может представлять собой глубокую сеть, то есть модель машинного обучения, составленную из нескольких уровней нелинейных операций.

5 Примерами глубоких сетей являются нейронные сети, включая сверточные нейронные сети, рекуррентные нейронные сети с одним или более скрытыми слоями и полносвязанные нейронные сети.

[0028] Сверточная нейронная сеть содержит архитектуры, которые могут обеспечить эффективную идентификацию текстовых полей. Сверточные нейронные сети могут  
10 содержать несколько сверточных и субдискретизирующих слоев, которые применяют фильтры к частям изображения документа для обнаружения определенных признаков. Таким образом, сверточная нейронная сеть включает операцию свертки, которая поэлементно умножает каждый фрагмент изображения на фильтры (например, матрицы) и суммирует результаты в аналогичной позиции выходного изображения (пример  
15 приведен на Фиг. 7).

[0029] Как отмечено выше, набор моделей машинного обучения 114 может быть обучен для определения типа поля с наибольшей уверенностью для поля 141 на изображении документа 140 с использованием данных обучения, как описано ниже. После обучения набора моделей машинного обучения 114, набор моделей машинного  
20 обучения 114 может быть передан в механизм идентификации текстового поля 112 для анализа новых изображений текста. Например, механизм идентификации текстового поля 112 может вводить анализируемое изображение документа 140 в набор моделей машинного обучения 114. Механизм идентификации текстового поля 112 может получать из набора обученных моделей машинного обучения 114 один или более выходных  
25 данных. Выходные данные являются оценкой качества одной или нескольких гипотез для типа поля для поля 141 (например, указатель на то, является ли гипотеза правильной).

[0030] Хранилище 120 представляет собой постоянную память, которая в состоянии сохранять изображения документов 140, а также структуры данных для разметки,  
30 организации и индексации изображений документов 140. Хранилище 120 может располагаться на одном или более запоминающих устройствах, таких как основное запоминающее устройство, магнитные или оптические запоминающие устройства на основе дисков, лент или твердотельных накопителей, NAS, SAN и т.д. Несмотря на то, что хранилище изображено отдельно от вычислительного устройства 110, в одной из  
35 реализаций изобретения хранилище 120 может быть частью вычислительного устройства 110. В некоторых вариантах реализации хранилище 120 может представлять собой подключенный к сети файловый сервер, в то время как в других вариантах реализации изобретения хранилище 120 может представлять собой какой-либо другой тип энергонезависимого запоминающего устройства, такой как объектно-ориентированная  
40 база данных, реляционная база данных и т.д., которая может находиться на сервере или на одной или более различных машинах, подключенных к нему через сеть 130.

[0031] В одном из вариантов реализации изобретения механизм идентификации текстового поля 112 начинает определение полей на изображении документа 140, создавая одну или более гипотез типа поля для поля 141. Чтобы определить одну или  
45 более гипотез, механизм идентификации текстового поля 112 может выполнять поиск по регулярным выражениям для определения типа данных, присутствующих в поле 141, или может применять шаблон к изображению документа 140 для определения ожидаемого типа поля, связанного с положением поля 141 на изображении документа

140. Сортировка гипотез, основанная на качестве, может быть выполнена, например, в тех случаях, если необходимо различать поля, содержащие похожие данные на чеках. В качестве примеров полей с похожими данными, которые могут быть выделены на чеках, можно привести следующие.

- 5 1. Денежная сумма: итого, сдача, оплата кредитной картой, скидка.
2. Денежные суммы в рамках позиций (вариант 1): цена товара; скидка; цена, включая скидки.
3. Денежные суммы в пределах позиций (вариант 2): цена за единицу и общая стоимость позиции.
- 10 4. Телефон / факс / телефон горячей линии.
5. Номер кредитной карты, номер дисконтной карты, номер подарочной карты или цифры со звездочками, которые не являются номером карты.
6. Почтовый индекс и номер дома в американских чеках.
7. Дата транзакции на чеке, дата, с которой вы можете вернуть товар, дата окончания
- 15 какого-либо действия, дата въезда на стоянку или выезда со стоянки и т.д.

[0032] На Фиг. 2А показано изображение чека 200, на котором имеются похожие типы данных (то есть похожие поля). Например, чек 200 содержит несколько денежных сумм для следующих позиций (Промежуточный итог 220, Итого 222, Дебетовая карта 224) или нескольких денежных сумм в пределах одной позиции (см. Фиг. 2В, иллюстрирующую фрагмент чека 200, соответствующий одной из позиций 230, где 232 - цена за единицу Цукини, 234 - общая стоимость продукта Цукини). Как описано более

20 подробно ниже, механизм идентификации текстового поля 112 позволяет отличать друг от друга эти поля и соответствующие значения.

[0033] На Фиг. 3 приведена блок-схема, иллюстрирующая способ идентификации поля в соответствии с одним или более вариантами реализации настоящего изобретения.

25 Способ 300 также может быть реализован при помощи вычислительной логики, содержащей аппаратное обеспечение (например, электронные схемы, специализированные логические схемы, программируемую логику, микрокод и т.п.), программное обеспечение (например, команды, выполняемые на обрабатывающем

30 устройстве для выполнения аппаратной имитации) или их сочетания. В одном из вариантов реализации изобретения способ 300 может выполняться вычислительным устройством 110, содержащим механизм генерации гипотез 111 и механизм идентификации текстового поля 112, как показано на Фиг. 1.

[0034] Как показано на Фиг. 3, на шаге 310 способ 300 получает одну или более

35 гипотез для типа поля первого текстового поля, присутствующего на изображении документа. В одном из вариантов реализации изобретения механизм идентификации текстового поля 112 может принимать запрос на выполнение идентификации поля на изображении документа, такого как изображение документа 200. Запрос может быть получен от пользователя вычислительного устройства 110, от пользователя клиентского

40 устройства, соединенного с вычислительным устройством 110 через сеть 130, или от какого-либо другого источника запроса.

[0035] В одном из вариантов реализации изобретения запрос включает в себя одну или более гипотез, созданных механизмом генерации гипотез 111 относительно типа поля для одного или более полей на изображении документа 140. Гипотезы могут

45 представлять собой исходное предположение или предсказание типа поля, выполненного с использованием в вычислительном отношении быстрых и дешевых техник. В качестве примера для генерации исходных гипотез, механизм генерации гипотез 111 может использовать простой механизм поиска полей по регулярными выражениями. Поиск

регулярного выражения может использоваться, чтобы отличить разные типы данных в чеке. Например, чтобы отличить денежные суммы от телефонных номеров, но это не поможет различить другие типы более похожих данных (например, различные типы денежных сумм, такие как итого, сдача, оплата по банковской карте, применяемая скидка и т.д.). В дополнение к регулярным выражениям, механизм генерации гипотез 111 может использовать для идентификации разных полей в чеке шаблоны. Шаблоны могут хранить информацию о структуре чека конкретного поставщика, включая ожидаемый тип поля, связанный с расположением поля в чеке. Механизм идентификации текстового поля 112 может сохранять принятую одну или более гипотез в хранилище 120.

[0036] На шаге 320 способ 300 создает трехмерную матрицу признаков, представляющую часть изображения, содержащую первое поле и связанный локальный контекст. В одном из вариантов реализации изобретения механизм идентификации текстового поля 112 выполняет ряд операций обработки изображения документа 200 для извлечения ряда признаков для ввода в модели машинного обучения 114. Например, первое измерение матрицы может быть измерением высоты, представляющим собой относительное положение вдоль оси Y (например, заданной строки), второе измерение матрицы может быть измерением ширины, представляющим собой относительное положение в указанной строке вдоль оси X (например, конкретной ячейки), а третье измерение матрицы может быть вектором признаков, представляющим собой значения признаков, извлеченных из позиции X-Y на изображении документа 200 и размещенных в определенном порядке. Обученные модули машинного обучения 114 могут использовать трехмерную матрицу признаков, представляющую часть изображения, содержащую первое поле и его локальный контекст, для идентификации и классификации типа поля любого поля текста, присутствующего на этой части изображения. Дополнительные подробности об обнаружении признаков, обработке изображений и генерации трехмерной матрицы признаков представлены ниже со ссылкой на Фиг. 4-6.

[0037] На шаге 330 способ 300 предоставляет трехмерную матрицу признаков в качестве входных данных в одну или более обученных моделей машинного обучения 114. В одном из вариантов реализации изобретения набор моделей машинного обучения 114 может быть составлен, например, из одного уровня линейных или нелинейных операций, таких как SVM или глубокая сеть (то есть, модель машинного обучения, составленная из нескольких уровней нелинейных операций), например, сверточная нейронная сеть. В одном из вариантов реализации изобретения сверточная нейронная сеть обучается с использованием обучающей выборки данных, содержащей примеры изображений документов, содержащих одно или более полей в качестве входных данных для обучения, и один или более идентификаторов типа поля, который правильно соответствует одному или более полям в качестве целевых выходных данных. Обучение может привести к оптимальной топологии сети. В одном из вариантов реализации изобретения слои сети могут содержать первый сверточный слой с окном фильтра  $1 \times 1$ . Одна ячейка матрицы признаков, сформированная выше (то есть значения признаков, соответствующие определенному положению  $x$  и  $y$ ), может считываться и подаваться на вход 20 нейронам. В одном из вариантов реализации изобретения может быть приблизительно 100 признаков, количество которых уменьшается до 20 признаков на выходе из первого сверточного слоя. Внутри каждой строки может быть еще один сверточный слой с окном фильтра  $1 \times 10$ . Таким образом, сеть может распространять (то есть извлекать) информацию из строки в местоположении. То есть, если есть какой-нибудь признак, сеть может определить не только, находится ли он в определенной

ячейке или нет, но и находится ли он также в соседних ячейках. Таким образом, сеть может получать признаки, учитывающие небольшой локальный контекст. Наконец, может быть полносвязанный слой (например, квадратная свертка 3×3). Количество нейронов в этом слое может зависеть от задачи, которая должна быть решена сетью.

5 [0038] На шаге 340 способ 300 получает результат из обученной модели машинного обучения, содержащий оценку качества одной или нескольких гипотез. Эта оценка качества одной или более гипотез, содержит, по меньшей мере, одно из: указание на то, что первая гипотеза из одной или более гипотез является предпочтительной гипотезой из множества гипотез, или значение уверенности, связанное с одной или более  
10 гипотезами. Если требуется отсортировать гипотезы по качеству (т.е. используется сценарий различия типа денежной суммы), то выходной слой может иметь несколько нейронов (например, по одному для каждого типа денежной суммы). Выход каждого нейрона может быть числом, которое характеризует оценку качества того, что рассматриваемые данные относятся к определенному классу (т.е. к типу поля). Если  
15 требуется только уверенность в том, что данные принадлежат определенному полю (т.е. указание на то, относится ли первое поле к данному типу поля: «да» или «нет»), выходной слой может включать один нейрон, который дает число, указывающее на уверенность в том, что данные соответствуют полю. Для разных полей топология может незначительно отличаться в зависимости от количества и качества данных,  
20 доступных для обучения. На Фиг. 7 показан один пример топологии сети для оценки уверенности гипотезы поля на чеке.

[0039] На Фиг. 7 показана топология сети для оценки уверенности гипотезы типа поля на изображении документа в соответствии с одним или более вариантами реализации настоящего изобретения. В одном из вариантов реализации изобретения  
25 топология сети представляет собой сверточную нейронную сеть, которая является частью набора моделей машинного обучения 114. Сверточная нейронная сеть содержит операцию свертки, которая может осуществлять умножение каждой позиции изображения на один или более фильтров (например, матриц свертки), как описано выше, поэлементно, с суммированием результата и его записью в аналогичной позиции  
30 выходного изображения. Сверточная нейронная сеть содержит входной слой и несколько сверточных и субдискретизирующих слоев. Например, сверточная нейронная сеть может включать в себя первый слой 702, имеющий тип входного слоя, второй слой 704, имеющий тип сверточного слоя, третий слой 706, имеющий тип сверточного слоя, четвертый слой 708, имеющий тип сверточного слоя, пятый слой 710, имеющий тип  
35 «MaxPooling» слоя, шестой слой 712, имеющий тип «Dropout» слоя, седьмой слой 714, имеющий тип «Flatten» слоя, восьмой слой 716, имеющий тип «Dense» слоя, девятый слой 718, имеющий тип «Dropout» слоя, десятый слой 720, имеющий тип «Dense» слоя, одиннадцатый слой 722, имеющий тип «Dropout» слоя, двенадцатый слой 724, имеющий тип «Dense» слоя, и тринадцатый слой 726, имеющий тип «Dense» слоя.

40 [0040] Обращаясь снова к Фиг. 3, на шаге 350 способ 300 выдает результаты поиска поля и указатель уверенности результатов.

[0041] На Фиг. 4 показана блок-схема, иллюстрирующая способ обработки изображения документа в соответствии с одним или более вариантами реализации настоящего изобретения. Способ 400 также может быть реализован при помощи  
45 вычислительной логики, содержащей аппаратное обеспечение (например, электронные схемы, специализированные логические схемы, программируемую логику, микрокод и т.п.), программное обеспечение (например, команды, выполняемые на обрабатывающем устройстве для выполнения аппаратной имитации) или их сочетания.

В одном из вариантов реализации изобретения способ 400 может выполняться механизмом идентификации текстового поля 112, как показано на Фиг. 1.

[0042] Как показано на Фиг. 4, на шаге 410 способ 400 определяет множество горизонтальных строк текста, присутствующих на изображении, причем одна строка из множества горизонтальных линий содержит первое поле. В одном из вариантов реализации изобретения, механизм идентификации текстового поля 112 может преобразовывать изображение, чтобы сделать все строки текста горизонтальными.

[0043] На шаге 420 способ 400 определяет систему координат для множества горизонтальных линий. В одном из вариантов реализации изобретения для задания системы координат механизм идентификации текстового поля 112 может находить на изображении левый и правый края документа, связывать первое значение с первым положением на пересечении левого края и, по меньшей мере, с одной из множества горизонтальных строк, а также связывать второе значение со вторым положением на пересечении правого края и, по меньшей мере, с одной из множества горизонтальных строк. Как показано на Фиг. 5, для каждой строки 502-510 механизм идентификации текстового поля 112 определяет систему координат. Пересечение левой границы чека 520 со строкой 506 обозначается как 0 (530), а пересечение правой границы чека 522 с линией 506 обозначается как 1 (532). Таким образом, все слова и символы, составляющие строку 506, будут расположены между 0 и 1 в определенной системе координат.

[0044] На шаге 430 способ 400 сдвигает систему координат по горизонтали на основании положения первого поля изображения, чтобы сформировать смещенную систему координат, причем трехмерная матрица признаков основана на смещенной системе координат. Чтобы сдвинуть систему координат по горизонтали, механизм идентификации текстового поля 112 сдвигает первое значение в положение первого поля изображения. Механизм идентификации текстового поля 112 может сдвигать систему координат по горизонтали таким образом, чтобы классифицируемые данные находились в середине соответствующей системы координат. Как далее показано на Фиг. 5, данные 540, подлежащие уточнению (т.е. для которых должна быть получена уверенность гипотезы) в исходной системе координат соответствующей строки, начинаются в точке с координатой 0,7 и заканчиваются в точке с координатой 0,8. Механизм идентификации текстового поля 112 преобразует заданную систему координат в другую систему координат, для которой координата 0,7 станет 0, а координата 0,8 станет 0,1. Новая система координат может быть расширена до интервала от -1 (550) до 1 (552). Аналогичное смещение выполняется для всех других строк (т.е. для всех строк точки с координатой 0,7 станут 0). Таким образом, весь чек будет вписываться в новую систему координат, где бы ни находилась интересующее поле, а само поле 540 будет находиться в центре новой системы координат. Такое смещение предоставит обученную модель машинного обучения 114 с более простой топологией. В одном из вариантов реализации изобретения трехмерная матрица признаков основана на этой смещенной системе координат.

[0045] На шаге 440 способ 400 кадрирует изображение для формирования кадрированного изображения, содержащего заданного количества строк выше и ниже одной из множества горизонтальных строк, которая содержит первое поле. В одном из вариантов реализации изобретения механизм идентификации текстового поля 112 кадрирует изображение, ограничивая его до 3-5 строк выше интересующей информации (строки) и того же количества строк ниже интересующей информации (строки). Это кадрирование основывается на предположении, что тип поля зависит только от локального контекста. В общем случае можно отправить все изображение чека на вход

сети, но обычно информация, расположенная далеко от данных, представляющих интерес, мало влияет на тип поля. В одном из вариантов реализации изобретения сеть принимает матрицу признаков фиксированного размера. Следовательно, механизм идентификации текстового поля 112 может фиксировать количество строк (то есть высоту матрицы). Если изображение кадрируется, чтобы получить 5 строк до и после интересующих данных, то высота матрицы признаков, поступающих на вход сети, составит 11.

[0046] На шаге 450 способ 400 разбивает кадрированное изображение на множество ячеек. В одном из вариантов реализации изобретения механизм идентификации текстового поля 112 разбивает результирующий прямоугольник на несколько частей по вертикали с интервалом, немного меньшим ширины символа (например, 80-100 частей). При этом данные разбиваются по ячейкам. В одном варианте осуществления ширина матрицы признаков также может иметь фиксированный размер. Так как ширина чеков может быть произвольной, с переменным числом символов в строках, механизм идентификации текстового поля 112 может разделить весь интервал от 1 до -1 на 80-100 частей одинакового размера.

[0047] На шаге 460 способ 400 вычисляет множество признаков для каждого из множества ячеек, причем множество признаков содержит информацию, относящуюся к графическим элементам, представляющим один или более символов, присутствующих в соответствующей ячейке. В одном из вариантов реализации изобретения, механизм идентификации текстового поля 112 использует информацию, полученную в результате оптического распознавания символов изображения чека и признаков, которые вычисляются по изображению (например, черная область, количество серий RLE). Признаки, которые вычисляются по изображению, являются скорее вспомогательными и могут использоваться для «нивелирования» ошибок идентификации. В общем, возможные признаки могут быть организованы в следующие классы. Среди этих признаков есть бинарные (например, есть буква (1) или нет (0)) и вещественные признаки.

[0048] Первый класс признаков содержит информацию об определенном распознанном символе (то есть, является ли этот символ специфичным Unicode, заглавная или строчная буква, класс символов (буква или цифра) и т.д.). Второй класс признаков содержит уверенность распознавания символов. Эти признаки сильно влияют на уверенность идентификации поля. Например, возможно, что мы почти уверены, что нашли поле в нужном месте, но также уверены, что мы распознали это поле с ошибками, поэтому мы не можем доверять значению поля, хотя оно и находится в правильном месте изображения. Третий класс признаков содержит признаки, которые характеризуют смысл слов, присутствующих на чеке. Такие признаки могут включать в себя словные эмбединги, присутствие в конкретном словаре и т.д. Эти признаки также характеризуют окружение поля, включая все другие слова в ближайшем окружении. Например, сеть может узнать, что если перед рассматриваемыми данными есть что-то о налогах и что-то о промежуточных итогах, то данные, вероятно, являются полем итоговой денежной суммы, даже если само слово ИТОГО не было распознано. Словным эмбедингам можно обучать по корпусам текстов или на текстах чек. Четвертый класс признаков содержит геометрические признаки, которые позволяют восстановить структуру чека. Эти признаки могут быть вычислены по изображению. Примеры геометрических признаков могут содержать подсчет количества черных пикселей, количество серий RLE, высоту строки и т.д. Кроме того, механизм идентификации текстового поля 112 может рассматривать признаки, связанные с шириной символов. В чеках некоторые буквы имеют двойной размер, т.е. занимают 2 моноширинные ячейки. На Фиг. 6 показаны

данные, в которых поле 602 содержит символы одинарной ширины, а поле 604 включает символы с удвоенным размером. Такие широкие буквы часто выделяют в чеке ключевые слова (например, слово ИТОГО). Даже если символ был распознан неправильно или вообще не распознан, информация о том, что этот символ является высоким или широким, может быть полезна для понимания того, что поблизости есть какое-то важное поле. Всего для каждой ячейки можно вычислить и сохранить около 100 признаков для ввода в сеть.

[0049] На шаге 470 способ 400 создает трехмерную матрицу признаков с использованием множества признаков как по меньшей мере одного компонента трехмерной матрицы признаков. Например, первое измерение матрицы может быть измерением высоты, представляющее собой относительное положение вдоль оси Y (например, заданной строки), второе измерение матрицы может быть измерением ширины, представляющее относительное положение в указанной строке вдоль оси X (например, конкретной ячейки), а третье измерение матрицы может быть вектором признаков, представляющее значения признаков, извлеченных из позиции X-Y на изображении документа 200 и размещенных в определенном порядке.

[0050] На Фиг. 8 приведен пример вычислительной системы 800, которая может выполнять один или более способов, описанных в настоящем документе, в соответствии с одним или более вариантами реализации настоящего изобретения. В одном из примеров вычислительная система 800 может соответствовать вычислительному устройству, способному выполнять функции механизм идентификации текстового поля 112, представленной на Фиг. 1. В другом примере вычислительная система 800 может соответствовать вычислительному устройству, способному выполнять функции механизма обучения 151, представленной на Фиг. 1. Эта вычислительная система 800 может быть подключена (например, по сети) к другим вычислительным системам в локальной сети, сети интранет, сети экстранет или сети Интернет. Данная вычислительная система 800 может выступать в качестве сервера в сетевой среде клиент-сервер. Эта вычислительная система 800 может представлять собой персональный компьютер (ПК), планшетный компьютер, телевизионную приставку (STB, set-top box), карманный персональный компьютер (PDA, Personal Digital Assistant), мобильный телефон, фотоаппарат, видеокамеру или любое устройство, способное выполнять набор команд (последовательно или иным способом), который определяется действиями этого устройства. Кроме того, несмотря на то, что показана система только с одним компьютером, термин «компьютер» также включает любой набор компьютеров, которые по отдельности или совместно выполняют набор команд (или несколько наборов команд) для выполнения одного или более любого из описанных здесь способов.

[0051] Пример вычислительной системы 800 включает устройство обработки 802, основное запоминающее устройство 804 (например, постоянное запоминающее устройство (ПЗУ), флэш-память, динамическое ОЗУ (DRAM, dynamic random access memory), например, синхронное DRAM (SDRAM, synchronous dynamic random access memory)), статическое запоминающее устройство 806 (например, флэш-память, статическое оперативное запоминающее устройство (SRAM, static random access memory)) и устройство хранения данных 818, которые взаимодействуют друг с другом по шине 830.

[0052] Устройство обработки 802 представляет собой одно или более устройств обработки общего назначения, например, микропроцессоров, центральных процессоров или аналогичных устройств. В частности, устройство обработки 802 может представлять

собой микропроцессор с полным набором команд (CISC, complex instruction set computing), микропроцессор с сокращенным набором команд (RISC, reduced instruction set computing), микропроцессор со сверхдлинным командным словом (VLIW, very long instruction word) или процессор, в котором реализованы другие наборов команд, или процессоры, в которых реализована комбинация наборов команд. Устройство обработки 802 также может представлять собой одно или более устройств обработки специального назначения, такое как специализированная интегральная схема (ASIC, application specific integrated circuit), программируемая пользователем вентильная матрица (FPGA, field programmable gate array), процессор цифровых сигналов (DSP, digital signal processor), сетевой процессор и т.д. Устройство обработки 802 реализовано с возможностью выполнения инструкций в целях выполнения рассматриваемых в этом документе операций и шагов.

[0053] Вычислительная система 800 может дополнительно включать устройство сопряжения с сетью 808. Вычислительная система 800 может также включать видеомонитор 810 (например, жидкокристаллический дисплей (LCD, liquid crystal display) или электронно-лучевую трубку (ЭЛТ)), устройство буквенно-цифрового ввода 812 (например, клавиатуру), устройство управления курсором 814 (например, мышь) и устройство для формирования сигналов 816 (например, динамик). В одном из иллюстративных примеров видео дисплей 810, устройство буквенно-цифрового ввода 812 и устройство управления курсором 814 могут быть объединены в один компонент или устройство (например, сенсорный жидкокристаллический дисплей).

[0054] Запоминающее устройство 818 может содержать машиночитаемый носитель 828, в котором хранятся инструкции 822 (например, механизм идентификации текстового поля 112 или механизм обучения 151), реализующие одну или более методологий или функций, описанных в данном документе. Инструкции 822 могут также находиться полностью или по меньшей мере частично в основном запоминающем устройстве 804 и (или) в устройстве обработки 802 во время их выполнения вычислительной системой 800, основным запоминающим устройством 804 и устройством обработки 802, также содержащим машиночитаемый носитель информации. Инструкции 822 могут дополнительно передаваться или приниматься по сети через устройство сопряжения с сетью 808.

[0055] Несмотря на то, что машиночитаемый носитель данных 828 показан в иллюстративных примерах как единичный носитель, термин «машиночитаемый носитель данных» следует понимать и как единичный носитель, и как несколько таких носителей (например, централизованная или распределенная база данных, и (или) связанные кэши и серверы), на которых хранится один или более наборов команд. Термин «машиночитаемый носитель данных» также следует понимать как включающий любой носитель, который может хранить, кодировать или переносить набор команд для выполнения машиной и который обеспечивает выполнение машиной любой одной или более методик настоящего изобретения. Соответственно, термин «машиночитаемый носитель данных» следует понимать как содержащий, среди прочего, устройства твердотельной памяти, оптические и магнитные носители.

[0056] Несмотря на то, что операции способов показаны и описаны в настоящем документе в определенном порядке, порядок выполнения операций каждого способа может быть изменен таким образом, чтобы некоторые операции могли выполняться в обратном порядке или чтобы некоторые операции могли выполняться, по крайней мере частично, одновременно с другими операциями. В некоторых вариантах реализации изобретения команды или подоперации различных операций могут выполняться с

перерывами и (или) попеременно.

[0057] Следует понимать, что приведенное выше описание носит иллюстративный, а не ограничительный характер. Различные другие варианты реализации станут очевидны специалистам в данной области техники после прочтения и понимания приведенного выше описания. Поэтому область применения изобретения должна определяться с учетом прилагаемой формулы изобретения, а также всех областей применения эквивалентных способов, которые покрывает формула изобретения.

[0058] В приведенном выше описании изложены многочисленные детали. Однако специалистам в данной области техники должно быть очевидно, что варианты реализации изобретения могут быть реализованы на практике и без этих конкретных деталей. В некоторых случаях хорошо известные структуры и устройства показаны в виде блок-схем, а не подробно, чтобы не усложнять описание настоящего изобретения.

[0059] Некоторые части представленных выше подробных описаний даны в виде алгоритмов и символического изображения операций с битами данных в компьютерной памяти. Такие описания и представления алгоритмов являются средством, используемым специалистами в области обработки данных, чтобы наиболее эффективно передавать сущность своей работы другим специалистам в данной области. Приведенный здесь (и в целом) алгоритм сформулирован как непротиворечивая последовательность шагов, ведущих к нужному результату. Эти шаги требуют физических манипуляций с физическими величинами. Обычно, хотя и не обязательно, эти величины принимают форму электрических или магнитных сигналов, которые можно хранить, передавать, комбинировать, сравнивать, и выполнять с ними другие манипуляции. Иногда удобно, прежде всего для обычного использования, описывать эти сигналы в виде битов, значений, элементов, символов, терминов, цифр и т.д.

[0060] Однако следует иметь в виду, что все эти и подобные термины должны быть связаны с соответствующими физическими величинами, и что они являются лишь удобными обозначениями, применяемыми к этим величинам. Если прямо не указано иное, как видно из последующего обсуждения, следует понимать, что во всем описании такие термины, как «прием» или «получение», «определение» или «обнаружение», «выбор», «хранение», «настройка» и т.п., относятся к действиям компьютерной системы или подобного электронного вычислительного устройства или к процессам в нем, причем такая система или устройство манипулирует данными и преобразует данные, представленные в виде физических (электронных) величин, в регистрах компьютерной системы и памяти в другие данные, также представленные в виде физических величин в памяти или регистрах компьютерной системы или в других подобных устройствах хранения, передачи или отображения информации.

[0061] Настоящее изобретение также относится к устройству для выполнения операций, описанных в настоящем документе. Такое устройство может быть специально сконструировано для требуемых целей, или оно может содержать универсальный компьютер, который избирательно активируется или дополнительно настраивается с помощью компьютерной программы, хранящейся в компьютере. Такая вычислительная программа может храниться на машиночитаемом носителе данных, включая, среди прочего, диски любого типа, в том числе гибкие диски, оптические диски, CD-ROM и магнитно-оптические диски, постоянные запоминающие устройства (ПЗУ), оперативные запоминающие устройства (ОЗУ), программируемые ПЗУ (EPROM), электрически стираемые ППЗУ (EEPROM), магнитные или оптические карты или любой тип носителя, пригодный для хранения электронных команд, каждый из которых соединен с шиной вычислительной системы.

[0062] Алгоритмы и изображения, приведенные в этом документе, не обязательно связаны с конкретными компьютерами или другими устройствами. Различные системы общего назначения могут использоваться с программами в соответствии с изложенной здесь информацией, возможно также признание целесообразным сконструировать более специализированные устройства для выполнения шагов способа. Структура разнообразных систем такого рода определяется в порядке, предусмотренном в описании. Кроме того, изложение вариантов реализации изобретения не предполагает ссылок на какие-либо конкретные языки программирования. Следует принимать во внимание, что для реализации принципов настоящего изобретения могут быть использованы различные языки программирования.

[0063] Варианты реализации настоящего изобретения могут быть представлены в виде вычислительного программного продукта или программы, которая может содержать машиночитаемый носитель данных с сохраненными на нем инструкциями, которые могут использоваться для программирования вычислительной системы (или других электронных устройств) в целях выполнения процесса в соответствии с сущностью изобретения. Машиночитаемый носитель данных включает механизмы хранения или передачи информации в машиночитаемой форме (например, компьютером). Например, машиночитаемый (считываемый компьютером) носитель данных содержит машиночитаемый (например, компьютером) носитель данных (например, постоянное запоминающее устройство (ПЗУ), оперативное запоминающее устройство (ОЗУ), накопитель на магнитных дисках, накопитель на оптическом носителе, устройства флэш-памяти и т.д.).

[0064] Слова «пример» или «примерный» используются здесь для обозначения использования в качестве примера, отдельного случая или иллюстрации. Любой вариант реализации или конструкция, описанные в настоящем документе как «пример», не должны обязательно рассматриваться как предпочтительные или преимущественные по сравнению с другими вариантами реализации или конструкциями. Слово «пример» лишь предполагает, что идея изобретения представляется конкретным образом. В этой заявке термин «или» предназначен для обозначения включающего «или», а не исключающего «или». Если не указано иное или не очевидно из контекста, то «X включает A или B» используется для обозначения любой из естественных включающих перестановок. То есть если X включает в себя A; X включает в себя B; или X включает A и B, то высказывание «X включает в себя A или B» является истинным в любом из указанных выше случаев. Кроме того, артикли «a» и «an», использованные в англоязычной версии этой заявки и в прилагаемой формуле изобретения, должны, как правило, означать «один или более», если иное не указано или из контекста не следует, что это относится к форме единственного числа. Использование терминов «вариант реализации» или «один вариант реализации» или «реализация» или «одна реализация» не означает одинаковый вариант реализации, если это не указано в явном виде. В описании термины «первый», «второй», «третий», «четвертый» и т.д. используются как метки для обозначения различных элементов и не обязательно имеют смысл порядка в соответствии с их числовым обозначением.

#### (57) Формула изобретения

1. Способ для идентификации текстовых полей на изображении документа, включающий:

получение устройством обработки изображения документа, содержащего одно или более текстовых полей;

генерация устройством обработки одной или более гипотез для типа поля первого текстового поля, содержащегося на изображении документа;

обработка изображения документа устройством обработки и создание трехмерной матрицы признаков, представляющей часть изображения, содержащую первое поле;

5 предоставление трехмерной матрицы признаков в качестве входных данных для обученной модели машинного обучения; и

получение выходных данных обученной модели машинного обучения, при этом выходные данные содержат оценку качества одной или более гипотез для типа поля первого текстового поля.

10 2. Способ по п. 1, в котором одна или более гипотез определяются с помощью поиска по регулярному выражению для определения типа данных, присутствующих в первом поле.

3. Способ по п. 1, в котором одна или более гипотез определяются с использованием шаблона, примененного к изображению, для определения ожидаемого типа поля, связанного с положением первого поля на изображении.

4. Способ по п. 1, дополнительно включающий:

определение множества горизонтальных текстовых строк, присутствующих на изображении, причем одна строка из множества горизонтальных строк содержит первое поле;

20 определение системы координат для множества горизонтальных строк; и смещение системы координат по горизонтали на основании положения первого поля на изображении для формирования смещенной системы координат.

5. Способ по п. 4, в котором определение системы координат содержит:

определение левого и правого краев документа на изображении;

25 связывание первого значения с первым положением на пересечении левого края и по меньшей мере одной строки из множества горизонтальных строк; и

связывание второго значения со вторым положением на пересечении правого края и по меньшей мере одной строки из множества горизонтальных строк;

30 в котором смещение системы координат по горизонтали содержит смещение первого значения в положение первого поля на изображении.

6. Способ по п. 4, в котором трехмерная матрица признаков основана на смещенной системе координат.

7. Способ по п. 4, дополнительно включающий

35 кадрирование изображения для формирования кадрированного изображения, содержащего заданное количество строк выше и ниже одной строки из множества горизонтальных строк, которая содержит первое поле.

8. Способ по п. 7, дополнительно включающий:

разбивку кадрированного изображения на множество ячеек; и

40 вычисление множества признаков для каждой ячейки из множества ячеек, причем множество признаков содержит по меньшей мере один компонент трехмерной матрицы признаков.

9. Способ по п. 8, в котором множество признаков содержит информацию, относящуюся к графическим элементам, представляющим один или более символов, присутствующих в соответствующей ячейке.

45 10. Способ по п. 1, в котором обученная модель машинного обучения содержит сверточную нейронную сеть.

11. Способ по п. 1, в котором оценка качества одной или более гипотез содержит по меньшей мере одно из: указание, что первая гипотеза из одной или более гипотез

является предпочтительной гипотезой из множества гипотез или значение уверенности, связанное с одной или более гипотезами.

12. Способ по п. 1, в котором обученная модель машинного обучения обучается с использованием обучающей выборки данных, содержащей примеры изображений документов, содержащих одно или более полей в качестве входных данных для обучения, и один или более идентификаторов типа поля, который правильно соответствует одному или более полям в качестве целевых выходных данных.

13. Система идентификации текстовых полей на изображении документа, содержащая: запоминающее устройство, в котором хранятся инструкции для выполнения вычислительной системы идентификации текстовых полей на изображении документа; устройство обработки, подключенное к указанному запоминающему устройству, выполненное с возможностью:

получения изображения документа, содержащего одно или более текстовых полей генерации одной или более гипотез для типа поля первого текстового поля, содержащегося на изображении документа;

обработки изображения документа устройством обработки и создания устройством обработки трехмерной матрицы признаков, представляющей часть изображения, содержащую первое поле;

предоставления трехмерной матрицы признаков в качестве входных данных для обученной модели машинного обучения; и

получения выходных данных из обученной модели машинного обучения, при этом выходные данные содержат оценку качества одной или более гипотез для типа поля первого текстового поля.

14. Система по п. 13, в которой устройство обработки дополнительно:

определяет множество горизонтальных текстовых строк, присутствующих на изображении, причем одна строка из множества горизонтальных строк содержит первое поле;

определяет систему координат для множества горизонтальных строк; и

сдвигает систему координат по горизонтали на основании положения первого поля изображения, чтобы сформировать смещенную систему координат, причем трехмерная матрица признаков основана на смещенной системе координат.

15. Система по п. 14, в которой устройство обработки дополнительно:

кадрирует изображение для формирования кадрированного изображения, содержащего заданное количество строк выше и ниже одной строки из множества горизонтальных строк, которая содержит первое поле;

разбивает кадрированное изображение на множество ячеек; и

вычисляет множество признаков для каждой из множества ячеек, причем множество признаков содержит информацию, относящуюся к графическим элементам, представляющим один или более символов, присутствующих в соответствующей ячейке, и содержит по меньшей мере один компонент трехмерной матрицы признаков.

16. Система по п. 13, в которой оценка качества одной или более гипотез содержит по меньшей мере одно из: указание, что первая гипотеза из одной или более гипотез является предпочтительной гипотезой из множества гипотез, или значение уверенности, связанное с одной или более гипотезами.

17. Постоянный машиночитаемый носитель данных, содержащий инструкции, побуждающие устройство обработки, взаимосвязанное с вычислительной системой, выполнять операции:

получение устройством обработки изображения документа, содержащего одно или

более текстовых полей;

генерация устройством обработки одной или более гипотез для типа поля первого текстового поля, содержащегося на изображении документа;

5 обработка изображения документа устройством обработки и создание трехмерной матрицы признаков, представляющей часть изображения, содержащую первое поле; предоставление трехмерной матрицы признаков в качестве входных данных для обученной модели машинного обучения; и

10 получение выходных данных обученной модели машинного обучения, при этом выходные данные содержат оценку качества одной или более гипотез для типа поля первого текстового поля.

18. Постоянный машиночитаемый носитель данных по п. 17, в котором устройство обработки дополнительно:

15 определяет множество горизонтальных текстовых строк, присутствующих на изображении, причем одна строка из множества горизонтальных строк содержит первое поле;

определяет систему координат для множества горизонтальных строк; и

сдвигает систему координат по горизонтали на основании положения первого поля изображения, чтобы сформировать смещенную систему координат, причем трехмерная матрица признаков основана на смещенной системе координат.

20 19. Постоянный машиночитаемый носитель данных по п. 18, в котором устройство обработки дополнительно:

кадрирует изображение для формирования кадрированного изображения, содержащего заданное количество строк выше и ниже одной строки из множества горизонтальных строк, которая содержит первое поле;

25 разбивает кадрированное изображение на множество ячеек; и

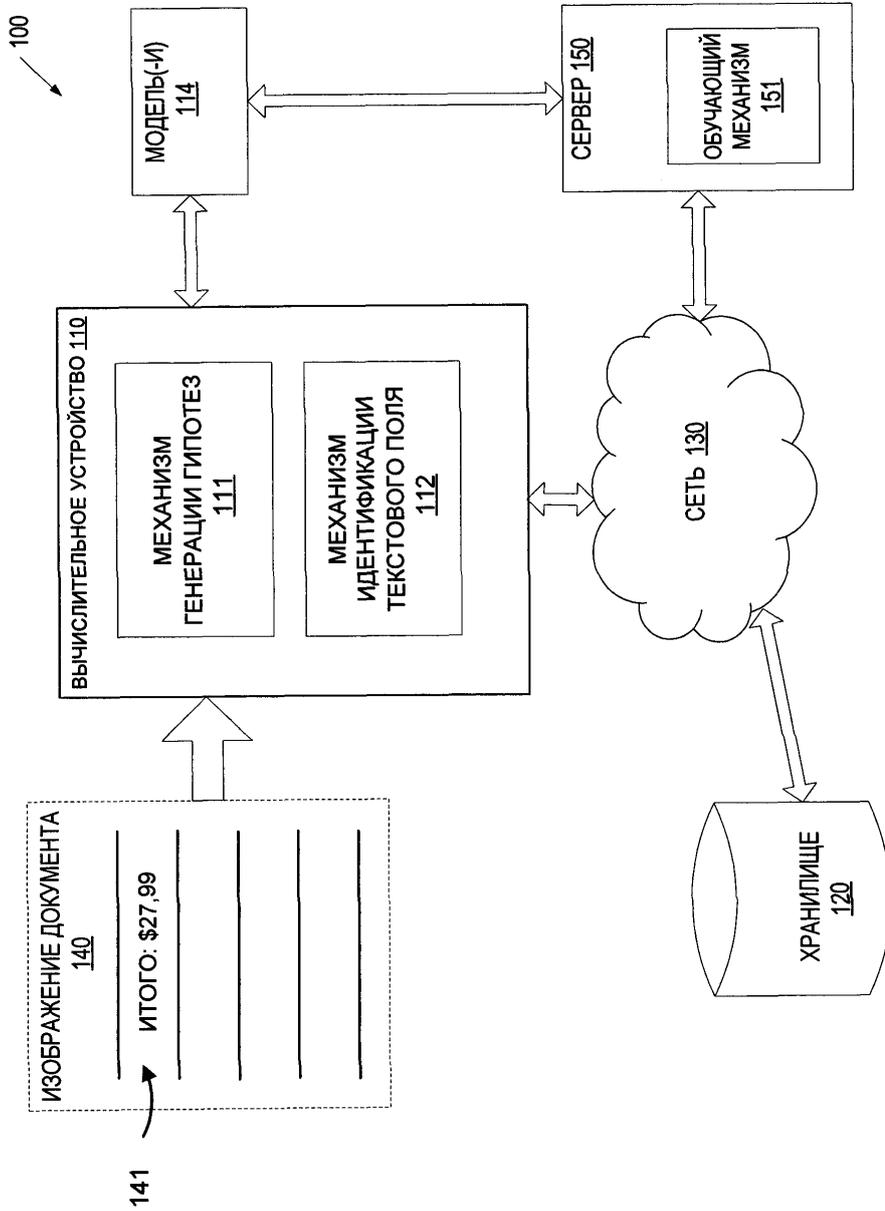
вычисляет множество признаков для каждой из множества ячеек, причем множество признаков содержит информацию, относящуюся к графическим элементам, представляющим один или более символов, присутствующих в соответствующей ячейке, и содержит по меньшей мере один компонент трехмерной матрицы признаков.

30 20. Постоянный машиночитаемый носитель данных по п. 17, в котором оценка качества одной или более гипотез содержит по меньшей мере одно из: указание, что первая гипотеза из одной или более гипотез является предпочтительной гипотезой из множества гипотез, или значение уверенности, связанное с одной или более гипотезами.

35

40

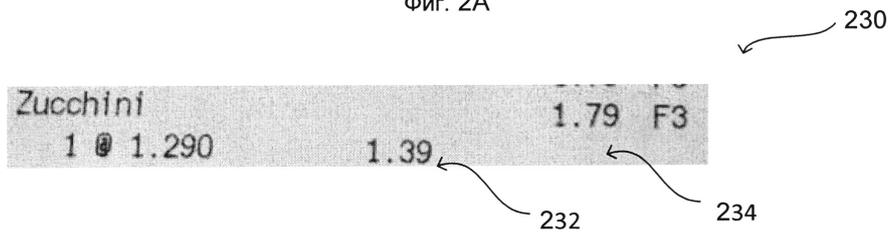
45



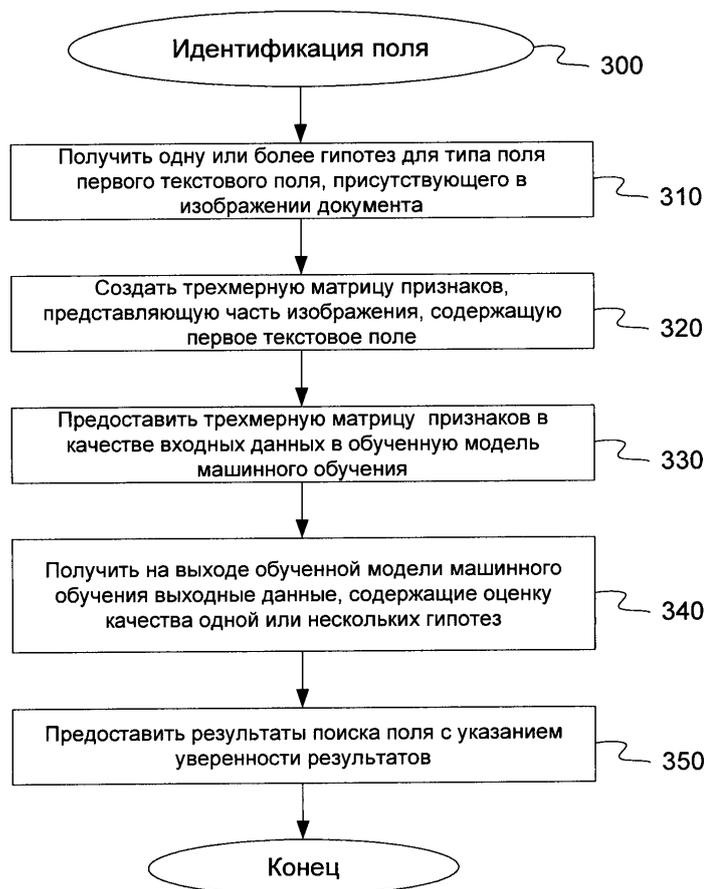
Фиг. 1



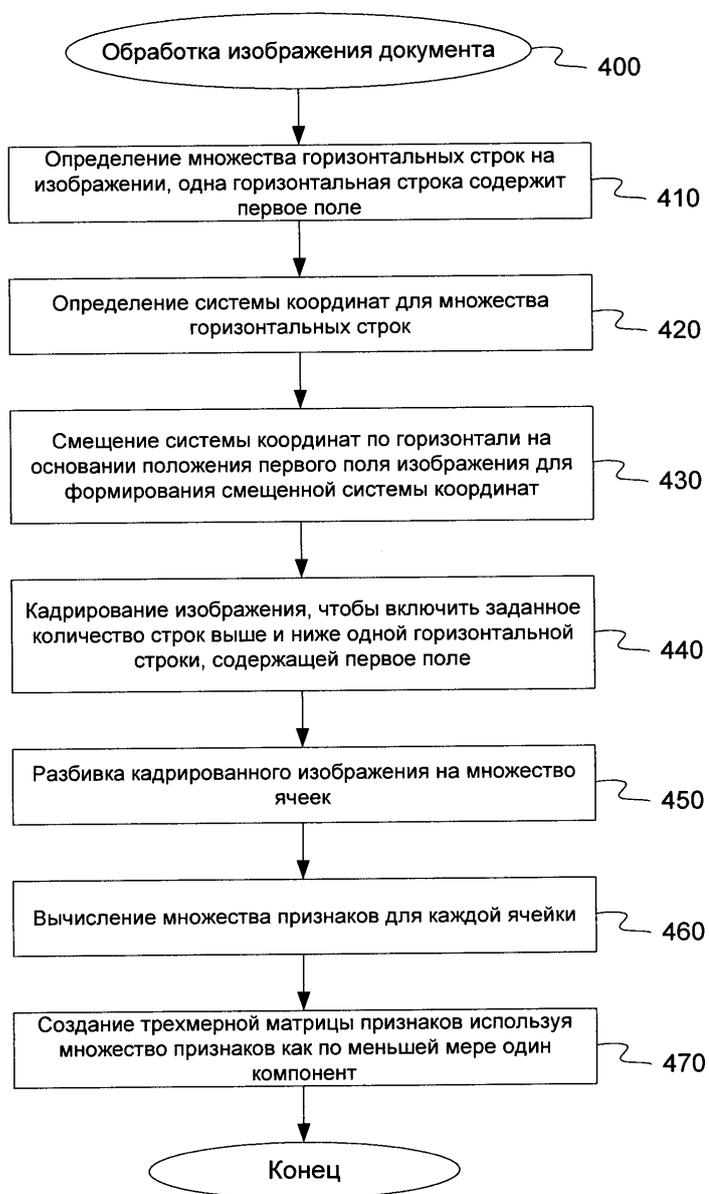
Фиг. 2А



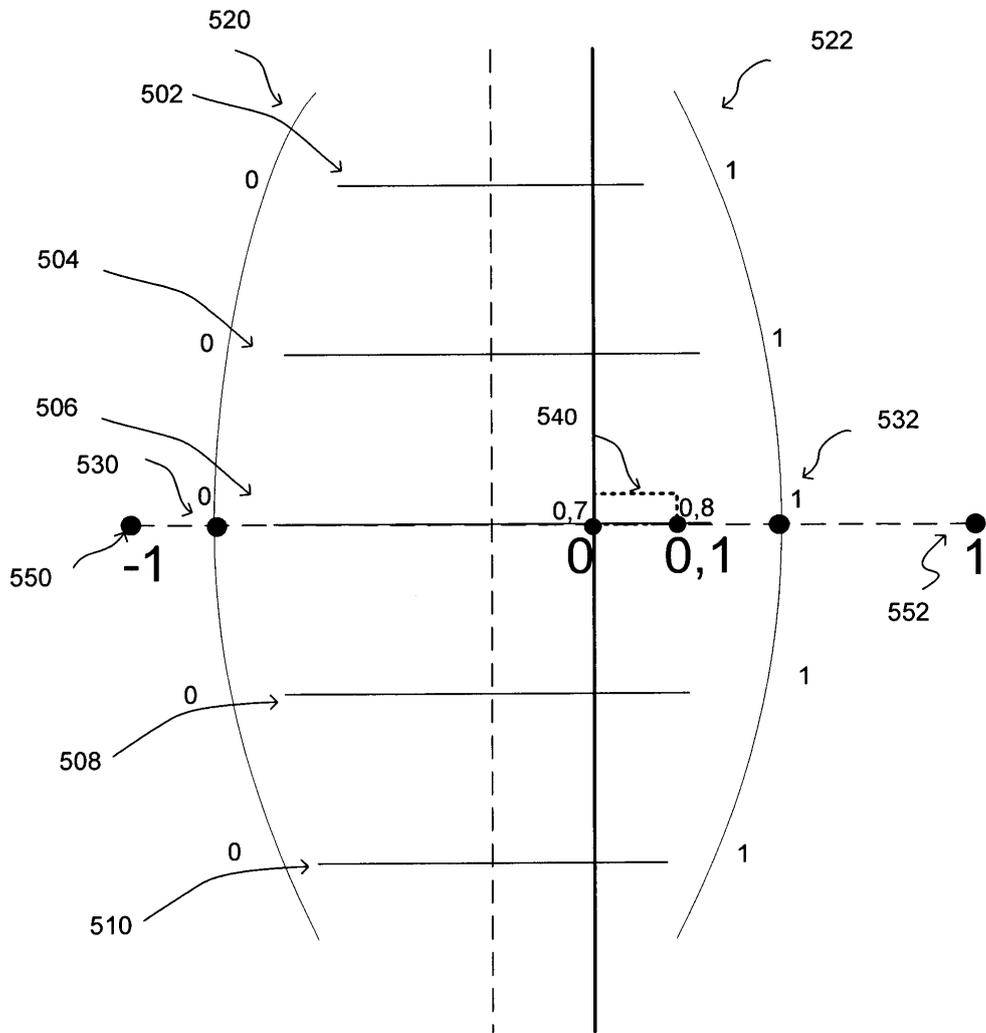
Фиг. 2В



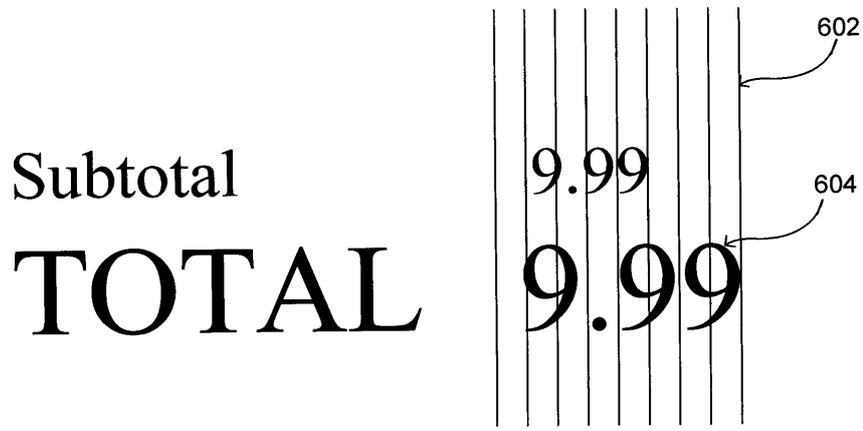
Фиг. 3



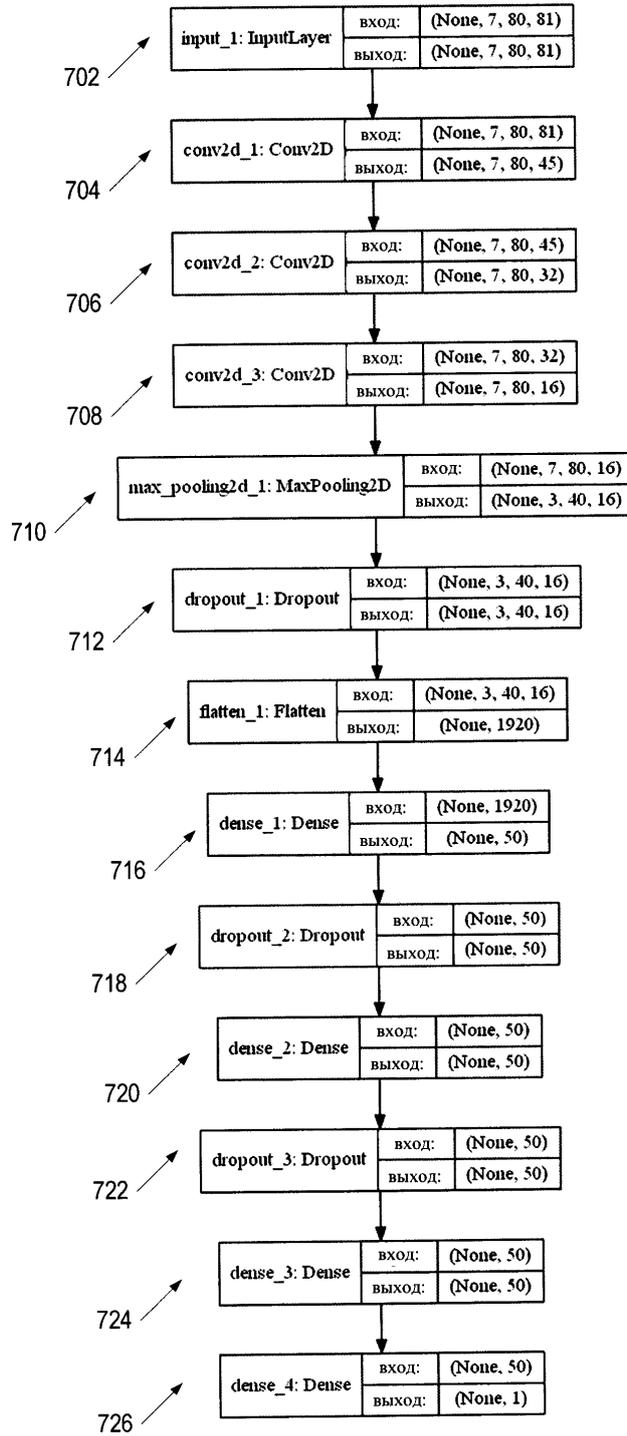
Фиг. 4



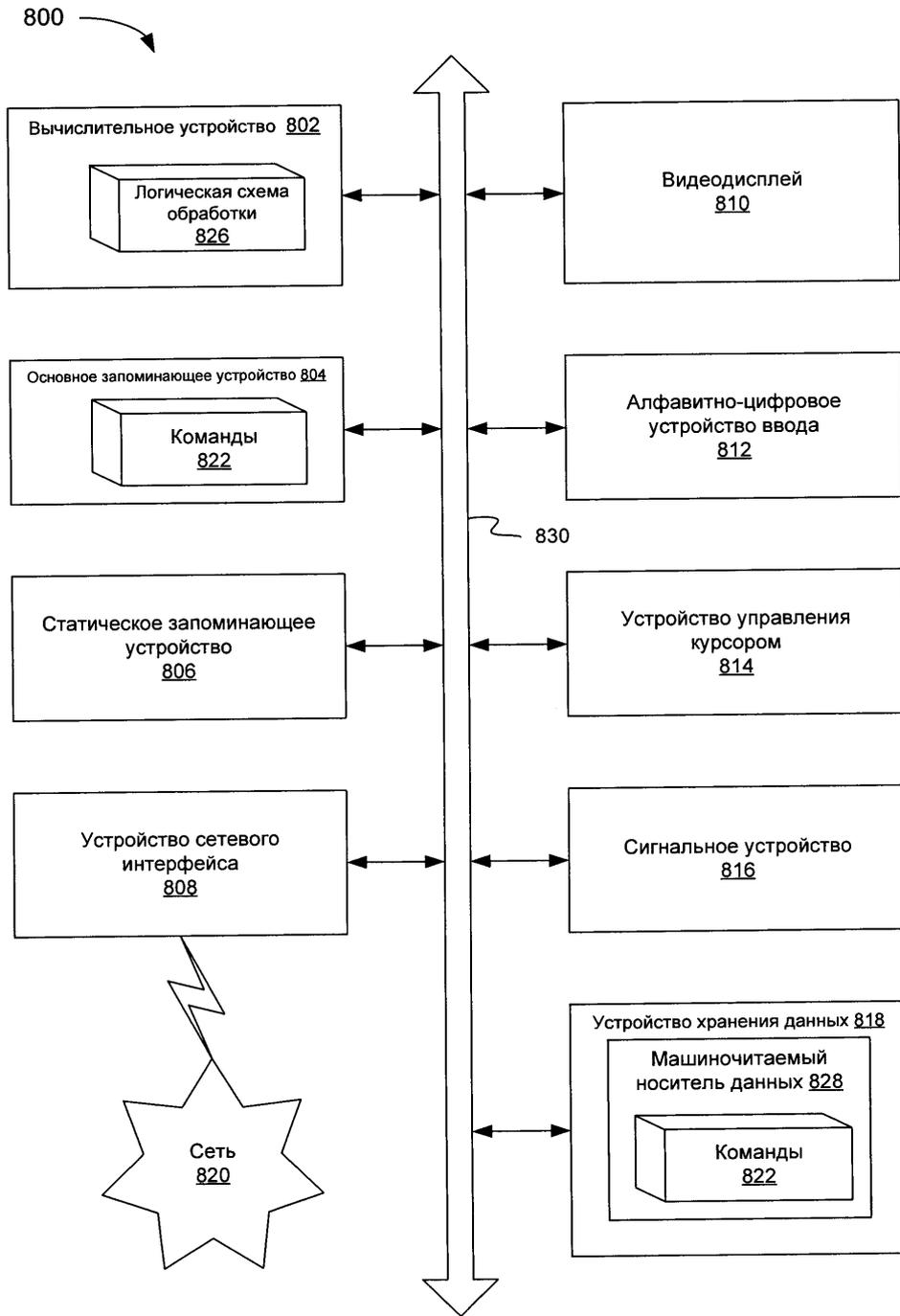
Фиг. 5



Фиг. 6



Фиг. 7



Фиг. 8