

(19)



**Евразийское
патентное
ведомство**

(21) **202293405** (13) **A1**

(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ**

(43) Дата публикации заявки
2023.08.24

(51) Int. Cl. **G06F 16/951** (2019.01)
G06F 17/00 (2019.01)
G06F 40/30 (2020.01)

(22) Дата подачи заявки
2022.12.20

(54) **СИСТЕМА И СПОСОБ СБОРА И ОБРАБОТКИ НОВОСТЕЙ В СЕТИ ИНТЕРНЕТ**

(31) **2022111786**

(32) **2022.04.29**

(33) **RU**

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

(72) Изобретатель:

**Шевцов Михаил Юрьевич, Козлов
Андрей Михайлович, Иванов
Александр Дмитриевич, Зубицкий
Павел Сергеевич, Малышев Илья
Александрович (RU)**

(74) Представитель:

Герасин Б.В. (RU)

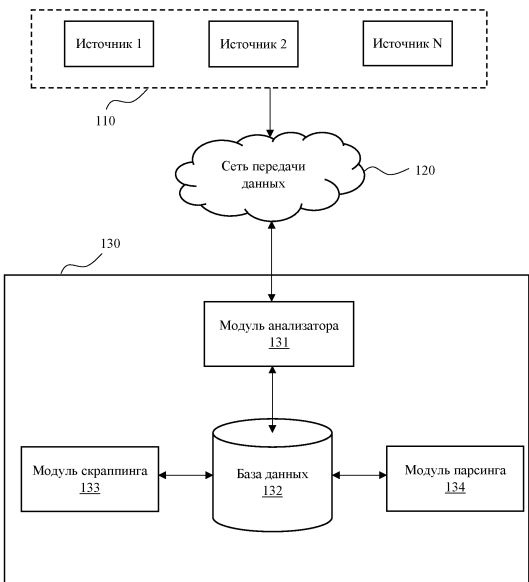
(57) Изобретение относится к области компьютерных технологий, в частности к автоматизированной системе для сбора данных в сети Интернет. Техническим результатом является повышение точности сбора новостных данных, за счет анализа новостных лент веб-сайтов на предмет наличия семантически связного текста в новостных источниках. Заявленный технический результат достигается за счет реализации системы сбора и обработки новостей в сети Интернет, содержащей: модуль анализатора, выполненный с возможностью поиска доменных имен в сети Интернет, содержащих новостные источники; анализа HTML-кода веб-страниц соответствующих доменных имен для выявления новостных лент; определение типа новостных лент и алгоритма обработки соответствующей ленты для извлечения ссылки на текстовую информацию новостного источника; передачу выявленных ссылок на новостные ленты, их тип и алгоритм обработки в базу данных; модуль скрапинга, выполненный с возможностью обработки данных, сохраненных в базе данных, с помощью которого осуществляется обработка сохраненных ссылок на новостные ленты с помощью применения алгоритма анализа разметки веб-ресурса, определенного модулем анализатора, при которой выполняется переход по ссылке на веб-ресурс, проверка ссылки на дубликацию с хранимой информацией в базе данных, и получение HTML-кода для последующей обработки текстовых данных; модуль парсинга, выполненный с возможностью получения HTML-кода от модуля скрапинга; извлечения текстовой информации из HTML-кода с помощью по меньшей мере двух алгоритмов сбора текстовых данных, каждый из которых, осуществляет выбор HTML-кода с наибольшим отношением символов, характеризующих связный текст новостного источника, к общему их количеству; обработку результатов извлечения каждого алгоритма моделью машинного обучения, причем модель выполнена с возможностью анализировать наличие характеристик, присущих источникам, не являющихся новостными источниками, при этом характеристики представляют собой по меньшей мере стоп-слова и спецсимволы; детектировать семантически связный текст, характеризующий новостной источник; сохранения извлеченного текста в базу данных.

A1

202293405

202293405

A1



СИСТЕМА И СПОСОБ СБОРА И ОБРАБОТКИ НОВОСТЕЙ В СЕТИ ИНТЕРНЕТ

ОБЛАСТЬ ТЕХНИКИ

[0001] Заявленное решение относится к области компьютерных технологий, в частности, к автоматизированной системе для сбора данных в сети Интернет.

УРОВЕНЬ ТЕХНИКИ

[0002] Автоматизированный сбор новостей в сети Интернет широко применяется на сегодняшний день. Часто применяются различные методы парсинга информации с новостных источников, позволяющие выгружать данные с веб-ресурсов для их последующей обработки.

[0003] В заявке US 20070198459 A1 (Current Assignee Accenture Global Services Ltd, 23.08.2007) раскрывается система для онлайн анализа источников информации, содержащая модуль сбора информации из сети и модуль аналитики, выполняющий анализ ретроспективного изменения данных в рамках анализируемого новостного топика.

[0004] Недостатками данного решения является отсутствие механизма проверки семантической связности текста, представленного в том или ином новостном источнике, что не позволяет проверить качество размещенной в сети информации, а также ее соответствие, как таковое, новостному источнику, а не иному типу данных, например, рекламному объявлению. В итоге такого рода решения позволяют осуществлять сбор данных без их предварительного анализа в части отношения к новостному источнику, в следствие чего снижается релевантность и качество собираемой информации.

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

[0005] Заявленная система позволяет решить техническую проблему в части повышения точности собираемой информации за счет выполнения проверки собираемой информации на предмет семантически связного текста, характеризующего новостной источник.

[0006] Техническим результатом является повышение точности сбора новостных данных, за счет анализа новостных лент веб-сайтов на предмет наличия семантически связного текста в новостных источниках.

[0007] Заявленный технический результат достигается за счет реализации системы сбора и обработки новостей в сети Интернет, содержащей:

модуль анализатора, выполненный с возможностью

поиска доменных имен в сети Интернет, содержащих новостные источники;
анализа HTML-кода веб страниц соответствующих доменных имен для выявления новостных лент;

определение типа новостных лент и алгоритма обработки соответствующей ленты для извлечения ссылки на текстовую информацию новостного источника;

передачу выявленных ссылок на новостные ленты, их тип и алгоритм обработки в базу данных;

модуль скраппинга, выполненный с возможностью обработки данных, сохраненных в базе данных, с помощью которого осуществляется

обработка сохраненных ссылок на новостные ленты с помощью применения алгоритма анализа разметки веб-ресурса, определенного модулем анализатора, при которой выполняется переход по ссылке на веб-ресурс, проверка ссылки на дубликацию с хранимой информацией в базе данных, и получение HTML-кода для последующей обработки текстовых данных;

модуль парсинга, выполненный с возможностью

получения HTML-кода от модуля скраппинга;

извлечение текстовой информации из HTML-кода с помощью по меньшей мере двух алгоритмов сбора текстовых данных, каждый из которых, осуществляет выбор HTML-ноды с наибольшим отношением символов, характеризующих связный текст новостного источника, к общему их количеству;

обработку результатов извлечения каждого алгоритма моделью машинного обучения, причем модель выполнена с возможностью

анализировать наличие характеристик, присущих источникам, не являющихся новостным источниками, при этом характеристики представляют собой по меньшей мере стоп-слова и спецсимволы;

детектировать семантически связный текст, характеризующий новостной источник;

сохранения извлеченного текста в базу данных.

[0008] В одном из частных примеров реализации анализ HTML-кода осуществляется для главной страницы веб-ресурса и для всех страниц 1-го уровня вложенности.

[0009] В другом частном примере реализации определяется наличие наличия ссылок, их количества и признаков совпадений по ключевым словам, соответствующих новостному источнику.

[0010] В другом частном примере реализации модуль скраппинга выполнен с возможностью анализа лент следующих типов:

- RSS - RSS, Atom, JSON стандарты;
- HTML страницы;
- HTML страницы, обрабатываемые с помощью XPath выражений.

[0011] Заявленный технический результат достигается также за счет выполнения способа сбора и обработки новостей в сети Интернет, выполняемого с помощью процессора и содержащего этапы, на которых:

выполняют поиск доменных имен в сети Интернет, содержащих новостные источники;

осуществляют анализ HTML-кода веб страниц соответствующих доменных имен для выявления новостных лент;

определяют тип новостных лент и алгоритм обработки соответствующей ленты для извлечения ссылки на текстовую информацию новостного источника;

передают выявленные ссылки на новостные ленты, их тип и алгоритм обработки в базу данных;

выполняют обработку данных, сохраненных в базе данных, в ходе которой обрабатывают сохраненные ссылки на новостные ленты с помощью алгоритма анализа разметки веб-ресурса, при этом выполняется переход по ссылке на веб-ресурс, проверка ссылки на дубликацию с хранимой информацией в базе данных, и получение HTML-кода для последующей обработки текстовых данных;

на основании полученного HTML-кода от модуля выполняют извлечение текстовой информации из HTML-кода с помощью по меньшей мере двух алгоритмов сбора текстовых данных, каждый из которых осуществляет выбор HTML-ноды с наибольшим отношением символов, характеризующих связный текст новостного источника, к общему их количеству;

обрабатывают результаты извлечения каждого алгоритма моделью машинного обучения, причем модель выполнена с возможностью

анализировать наличие характеристик, присущих источникам, не являющихся новостным источниками, при этом характеристики представляют собой по меньшей мере стоп-слова и спецсимволы;

детектировать семантически связный текст, характеризующий новостной источник;

сохранения извлеченного текста в базу данных.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

[0012] На Фиг. 1 представлена концептуальная схема заявленного решения.

[0013] На Фиг. 2 представлен пример HTML-кода, извлекаемого с ресурса модулем анализатора.

[0014] На Фиг. 3 представлен пример извлечения ссылок из HTML-кода.

[0015] На Фиг. 4 представлен пример записи ссылки на новостной источник в базе данных.

[0016] На Фиг. 5 представлен пример XPath выражения.

[0017] На Фиг. 6 представлен пример обработки HTML ленты.

[0018] На Фиг. 7 представлен пример извлеченного текста из новостного источника.

[0019] На Фиг. 8 представлена общая схема вычислительного устройства.

ОСУЩЕСТВЛЕНИЕ ИЗОБРЕТЕНИЯ

[0020] На Фиг. 1 представлена общая схема заявленной системы (130), выполняющей сбор информации с веб-сайтов, содержащих новостные источники (110). Система (130) может выполняться на базе единого вычислительного устройства, например, сервера, или представлять собой программно-аппаратный комплекс, в котором каждый из ее элементов расположен на отдельной вычислительной машине, связанный в рамках единого функционального обеспечения с другими элементами посредством информационной сети.

[0021] Система (130) содержит совокупность модулей, которые реализуют заданный функционал. Модули могут выполняться конструктивно в виде программно-аппаратных решений (например, система на чипе, микроконтроллеров и т.п.) или в виде программных модулей, функционирующих в рамках единого программного обеспечения, реализующего алгоритм работы системы (130) с помощью вычислительного устройства.

[0022] Сбор информации системой (130) из сети Интернет осуществляется посредством модуля анализатора (131), выполняющему подключение к веб-сайтам с новостными источниками (110) через информационную сеть (120).

[0023] Модуль анализатора (131) выполняет поиск доменных имен в сети Интернет, содержащих новостные источники (110). После подключения к источникам модуль (131) выполняет анализ HTML-кода веб-страниц соответствующих доменных имен на предмет выявления новостных лент. Анализ новостных источников выполняется с помощью анализа главной страницы веб-ресурса, а также всех страниц 1-го уровня вложенности. На Фиг. 2 представлен пример извлечения HTML-кода из источника (110) домена <https://press.sber.ru>.

[0024] Алгоритм обработки модуля (131) использует два типа алгоритмов `rssfinder` и `htmlfinder`, которые обеспечивают анализ веб-страниц и выявление ссылок на RSS-ленты или HTML-ленты. Пример определения ссылок на новостные ленты представлен на Фиг. 3. После выявления одной или нескольких новостных лент модуль (131) выполняет определение типа новостных лент и алгоритма обработки соответствующей ленты для извлечения ссылки на текстовую информацию новостного источника. Первым работает алгоритм `rssfinder`, т.к. RSS ленты проще в обработке, если по итогу работы `rssfinder` ничего не выявил, то активируется алгоритм `htmlfinder`. При этом возможны случаи, когда ссылка некорректная или источник не доступен (нет отклика от сервера), в таком случае тип ленты определяется в процессе выявления и зависит от того, какой алгоритм вернул значения, и самого факта получения ответов от сервера источника.

[0025] При работе модуля (131) также определяется наличие ссылок, их количества и признаков совпадений по ключевым словам, соответствующих новостному источнику, например, таким как: "rss", "feed", "news", "articles", "новости", "статьи", или исключающих (".png", ".pdf", паттерны: `.*login.*`, `.*\/([\^-\]*[-_][\^-\]*)+ $\$$` и др.). Выявленные ссылки на новостные ленты, а также их тип (HTML или RSS) и применимый алгоритм обработки для последующего извлечения ссылок на новости передаются в базу данных (132). Пример записи в базе данных представлен на Фиг. 4.

[0026] Сохраненная информация в базе данных (132) далее обрабатывается с помощью модулей скраппинга (133) и парсинга (134). Модуль скраппинга (133) обеспечивает обработку сохраненных ссылок на новостные ленты с помощью применения алгоритма анализа разметки веб-ресурса, определенного модулем анализатора (131), при которой выполняется переход по ссылке на веб-ресурс источника (110) для проверки ссылки на дубликацию с хранимой информацией в базе данных (132), а также получения HTML-кода

для последующей обработки текстовых данных модулем парсинга (134). Пример извлечения ссылок из HTML-кода представлен на Фиг. 3.

[0027] Модуль скраппинга (133) выполняет непрерывную работу и итеративно обрабатывает таблицу ссылок на ленты из базы данных (132). В асинхронном режиме в модуле (133) работает три цикла, поддерживающие обработку 3-х типов лент: RSS - RSS, Atom, JSON стандарты (1й тип); HTML - обычные HTML-страницы (2й тип); HTML-страницы, обрабатываемые с помощью XPath выражений (3й тип), для которых путь к ссылкам на новости конфигурируется вручную. Пример записи XPath выражения представлен на Фиг. 5.

[0028] Каждый из циклов обрабатывает часть соответствующих его алгоритму ссылок, при работе которых происходит обращение по ссылке к источнику (110) для анализа полученного HTML-кода для извлечения ссылок на новостные данные. Пример промежуточной обработки для HTML ленты представлен на Фиг. 6. Все полученные ссылки на новости проверяются на дублирование путем обращения к базе данных (132), если ссылка содержится в базе данных (132) - то она исключается из обработки, иначе - записывается в базу данных (132) и передается на дальнейшую обработку.

[0029] Модуль парсинга (134) обрабатывает полученный HTML-код от модуля скраппинга (133). В ходе работы модуля (134) осуществляется извлечение текстовой информации из HTML-кода с помощью по меньшей мере двух алгоритмов сбора текстовых данных, каждый из которых, осуществляет выбор HTML-ноды с наибольшим отношением символов, характеризующих связный текст новостного источника, к общему их количеству. Под HTML-нодой понимается иерархический узел HTML разметки, например, <head>, <body> и т.п.

[0030] Один из применяемых алгоритмов основан на измерении количества непробельных символов в HTML-ноде источника. Другой алгоритм анализирует HTML-ноды по количеству полезного текста, и выполняет извлечение текста из набравших больший вес нод. Путем тестирования этих алгоритмов на одном наборе данных была выявлена разность множеств качественных текстов. Алгоритмы работают параллельно и сравнивается оценка результатов моделью машинного обучения, например, нейросетевым алгоритмом, обученным на примерах новостных источников, представляющих собой эталонные новостные тексты. Модель машинного обучения, применяемая в рамках работы модуля парсинга (134) выполняет анализ наличия характеристик, присущих источникам, не являющихся новостными источниками. Такого рода характеристики, как правило, представляют собой стоп-слова и спецсимволы (например, номера телефонов, последовательность цифр, и т.п.). На основании обработки моделью получаемых

результатов работы указанных выше алгоритмов выполняется выявление наиболее семантически связного текста, что явно характеризует новостной источник. Полученный текст впоследствии сохраняется в базу данных (132) для последующего предоставления пользователю или передачи в автоматизированную систему подбора новостей по ключевым словам. Пример извлеченного текста представлен на Фиг. 7

[0031] Заявленная система (130) может быть реализована на базе единого вычислительного устройства (200), например, сервере. На Фиг. 8 представлен общий вид такого вычислительного устройства (200).

[0032] В общем случае вычислительное устройство (200) содержит объединенные общей шиной информационного обмена один или несколько процессоров (201), средства памяти, такие как ОЗУ (202) и ПЗУ (203), интерфейсы ввода/вывода (204), устройства ввода/вывода (205), и устройство для сетевого взаимодействия (206).

[0033] Процессор (201) (или несколько процессоров, многоядерный процессор) могут выбираться из ассортимента устройств, широко применяемых в текущее время, например, компаний Intel™, AMD™, Apple™, Samsung Exynos™, MediaTek™, Qualcomm Snapdragon™ и т.п. В качестве процессора (501) может также применяться графический процессор, например, Nvidia, AMD, Graphcore и пр.

[0034] ОЗУ (202) представляет собой оперативную память и предназначено для хранения исполняемых процессором (201) машиночитаемых инструкций для выполнения необходимых операций по логической обработке данных. ОЗУ (202), как правило, содержит исполняемые инструкции операционной системы и соответствующих программных компонент (приложения, программные модули и т.п.).

[0035] ПЗУ (203) 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.

[0036] Для организации работы компонентов устройства (200) и организации работы внешних подключаемых устройств применяются различные виды интерфейсов В/В (204). 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.

[0037] Для обеспечения взаимодействия пользователя с вычислительным устройством (500) применяются различные средства (205) 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.

[0038] Средство сетевого взаимодействия (206) обеспечивает передачу данных устройством (200) посредством внутренней или внешней вычислительной сети, например, Интранет, Интернет, ЛВС и т.п. В качестве одного или более средств (206) может использоваться, но не ограничиваться: 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.

[0039] Дополнительно могут применяться также средства спутниковой навигации в составе устройства (200), например, GPS, ГЛОНАСС, BeiDou, Galileo.

[0040] 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

1. Система сбора и обработки новостей в сети Интернет, содержащая:

модуль анализатора, выполненный с возможностью

поиска доменных имен в сети Интернет, содержащих новостные источники;
анализа HTML-кода веб страниц соответствующих доменных имен для выявления новостных лент;

определение типа новостных лент и алгоритма обработки соответствующей ленты для извлечения ссылки на текстовую информацию новостного источника;

передачу выявленных ссылок на новостные ленты, их тип и алгоритм обработки в базу данных;

модуль скраппинга, выполненный с возможностью обработки данных, сохраненных в базе данных, с помощью которого осуществляется

обработка сохраненных ссылок на новостные ленты с помощью применения алгоритма анализа разметки веб-ресурса, определенного модулем анализатора, при которой выполняется переход по ссылке на веб-ресурс, проверка ссылки на дубликацию с хранимой информацией в базе данных, и получение HTML-кода для последующей обработки текстовых данных;

модуль парсинга, выполненный с возможностью

получения HTML-кода от модуля скраппинга;

извлечения текстовой информации из HTML-кода с помощью по меньшей мере двух алгоритмов сбора текстовых данных, каждый из которых осуществляет выбор HTML-ноды с наибольшим отношением символов, характеризующих связный текст новостного источника, к общему их количеству;

обработку результатов извлечения каждого алгоритма моделью машинного обучения, причем модель выполнена с возможностью

анализировать наличие характеристик, присущих источникам, не являющихся новостным источниками, при этом характеристики представляют собой по меньшей мере стоп-слова и спецсимволы;

детектировать семантически связный текст, характеризующий новостной источник;

сохранения извлеченного текста в базу данных.

2. Система по п.1, характеризующая тем, что анализ HTML-кода осуществляется для главной страницы веб-ресурса и для всех страниц 1-го уровня вложенности.

3. Система по п.2, характеризующая тем, что определяется наличие наличия ссылок, их количества и признаков совпадений по ключевым словам, соответствующих новостному источнику.

4. Система по п.1, характеризующая тем, что модуль скраппинга выполнен с возможностью анализа лент следующих типов:

- RSS - RSS, Atom, JSON стандарты;
- HTML страницы;
- HTML страницы, обрабатываемые с помощью XPath выражений.

5. Способ сбора и обработки новостей в сети Интернет, выполняемый с помощью процессора и содержащий этапы, на которых:

выполняют поиск доменных имен в сети Интернет, содержащих новостные источники;

осуществляют анализ HTML-кода веб страниц соответствующих доменных имен для выявления новостных лент;

определяют тип новостных лент и алгоритм обработки соответствующей ленты для извлечения ссылки на текстовую информацию новостного источника;

передают выявленные ссылки на новостные ленты, их тип и алгоритм обработки в базу данных;

выполняют обработку данных, сохраненных в базе данных, в ходе которой обрабатывают сохраненные ссылки на новостные ленты с помощью алгоритма анализа разметки веб-ресурса, при этом выполняется переход по ссылке на веб-ресурс, проверка ссылки на дубликацию с хранимой информацией в базе данных, и получение HTML-кода для последующей обработки текстовых данных;

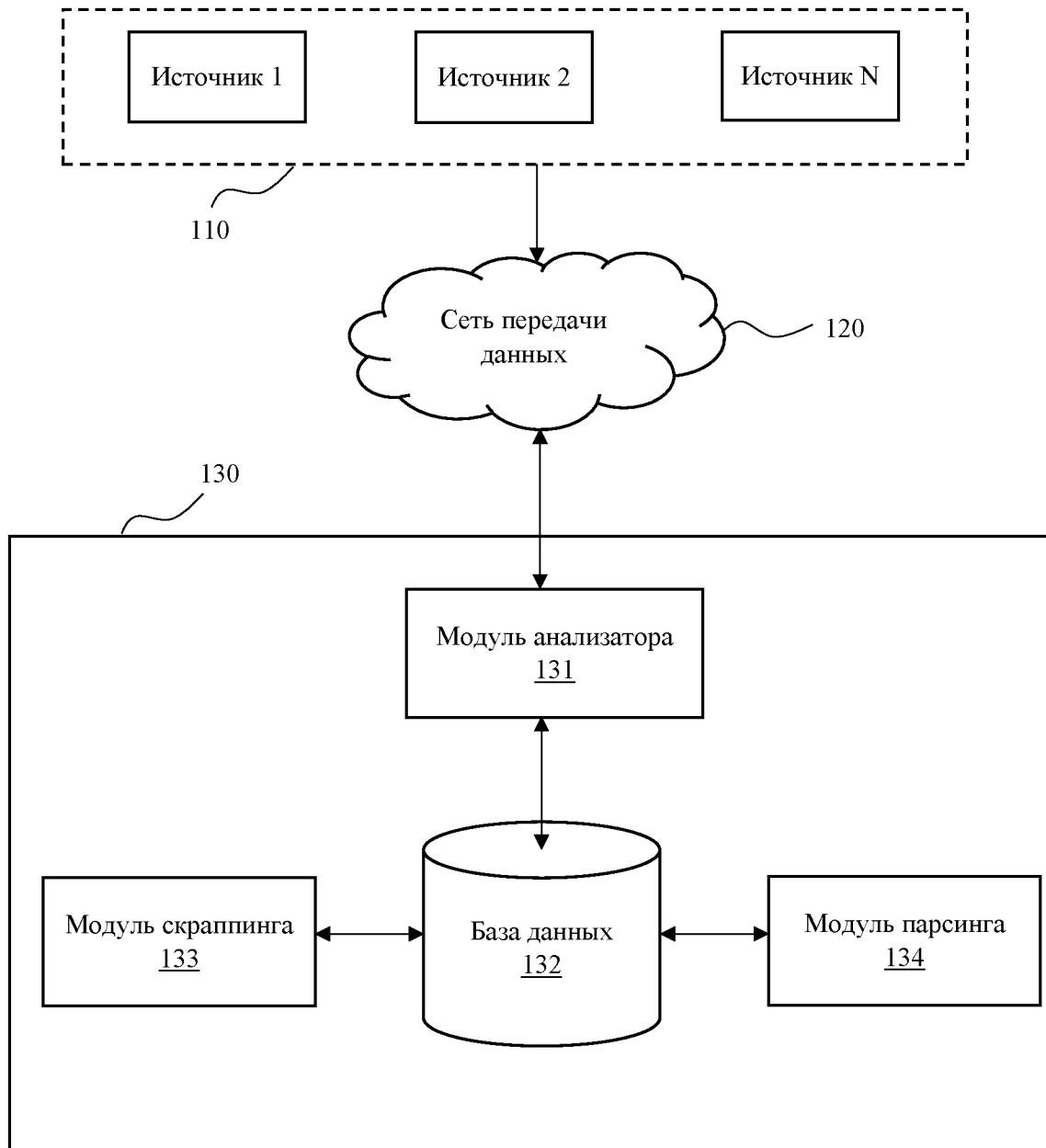
на основании полученного HTML-кода от модуля выполняют извлечение текстовой информации из HTML-кода с помощью по меньшей мере двух алгоритмов сбора текстовых данных, каждый из которых осуществляет выбор HTML-ноды с наибольшим отношением символов, характеризующих связный текст новостного источника, к общему их количеству;

обрабатывают результаты извлечения каждого алгоритма моделью машинного обучения, причем модель выполнена с возможностью

анализировать наличие характеристик, присущих источникам, не являющихся новостным источниками, при этом характеристики представляют собой по меньшей мере стоп-слова и спецсимволы;

детектировать семантически связный текст, характеризующий новостной источник;

сохранения извлеченного текста в базу данных.



Фиг. 1

```

<html lang="ru" style="--vh:9.84px;">
  <head>...</head>
  <body class="using-mouse">
    <div id="__next" == $0
      <header class="_2jUuA"> flex
        <div class="_33LNT_1VNPv"> flex
          <div class="_3sGVs"> flex
            <div class="_30q0q"> flex
              <div>...</div>
              <a href="/">...</a>
              <nav class="_1sm63">
                <a class="_1z6e4" href="https://www.sber.ru">Для меня</a> flex
                <a class="_1z6e4" href="https://www.sber.ru/business">Для бизнеса</a> flex
                <a class="_1z6e4_23Lxr" href="/">Медиа</a> flex
              </nav>
            </div>
          <div class="_2h94r">...</div> flex
        </div>
      </div>
    <div class="_33LNT_2WuNU">
      <div class="_3sGVs"> flex
        <nav class="_2-iPa">
          <a class="ZIw0Y w2CyQ" href="/">Главная</a>
          <a class="ZIw0Y" href="/publications">Новости</a>
          <div class="_1n-dZ ZIw0Y">
            <button class="_25Mle_25Mle" aria-label="Открыть список" type="button">...</button> flex
            <ul class="_3dz5S">...</ul>
          </div>
          <a class="ZIw0Y" href="/columns">Колонки</a>
          <a class="ZIw0Y" href="/topics/analitikaissledovaniia">Аналитика</a>
          <a class="ZIw0Y" href="/gallery">Фото и видео</a>
          <div class="_1n-dZ ZIw0Y">...</div>
        </nav>
        <div class="_2Ylyp">...</div>
      </div>
    </div>
  </header>

```

Фиг. 2

<https://press.sber.ru/publications>
<https://press.sber.ru/topics/finance>
<https://press.sber.ru/topics/ekosistema>
<https://press.sber.ru/topics/business>
<https://press.sber.ru/topics/security>
<https://press.sber.ru/topics/technology>
<https://press.sber.ru/topics/ekologichnost>
<https://press.sber.ru/topics/talents>
<https://press.sber.ru/topics/education>
<https://press.sber.ru/topics/health>
<https://press.sber.ru/topics/ofitsialnoe>
<https://press.sber.ru/topics/analitikaissledovaniia>

Фиг. 3

```

id : id_domain : id_link_source_type : url_link_source : is_active : dtm_inserted : xpath :
24876 48854 2 http://www.sberbank.com/ru/news true 2020-06-17 02:37:33.695261 <nil>

```

Фиг. 4

The screenshot shows a web application interface. On the left, there is a table with columns: ID, Доменная имя (Domain name), Тип (Type), URL, and Язык (Language). The table contains one row with ID 71817, domain name press.sber.ru, type news, URL http://press.sber.ru, and language Russ. Below the table are sections for 'Источники новостей' (News sources) with a link to 'https://press.sber.ru/publications', 'Хосты выделены' (Hosts are highlighted) with a link to 'ru@class-100.ru', and 'Комментарий к источнику' (Comment on source). On the right, there is a 'Результаты' (Results) section with a list of URLs and a 'Ссылки вкладки' (Link tabs) section.

Фиг. 5

The screenshot shows a web browser displaying the 'SBER PRESS' website. The address bar shows 'press.sber.ru/publications'. The page features two news articles. The first article is titled 'Ситидрайв запустил новую программу лояльности' (Citidrive launched a new loyalty program) and mentions that privileges will be available to all users. The second article is titled 'СберРешения получили аккредитацию IT-компаний' (SberResheniya received IT company accreditation) and mentions that the company has passed state accreditation and received IT status.

Фиг. 6

Результат

Сервис Ситидрайв запустил новую систему поощрения пользователей каршеринга. Вместо процентной скидки появится кэшбэк до 15%. В зависимости от уровня пользователя, накопленными бонусами можно списать от 25% до 60% стоимости на поездки и дополнительные опции.

В рамках обновлённой программы лояльности пользователям доступны пять уровней: «Автолюбитель», «Заводила», «Энтузиаст», «Человек машина», «Звезда дороги». Специально для новых пользователей также создан welcome-уровень «Исследователь» с повышенным кэшбэком на 30 дней с момента регистрации. Таким образом, привилегии теперь доступны каждому пользователю.

Переход на следующий уровень теперь зависит от километража: то есть, сколько километров пользователь проехал через сервис Ситидрайв, столько и учитывается в его прогрессе. Лимита по количеству поездок для перехода на более высокий уровень нет – например, после завершения длинной поездки есть возможность перейти со второго уровня сразу на четвёртый, минуя при этом третий. Также в новой программе лояльности не будет отрицательных уровней – накопленные километры не будут списываться.

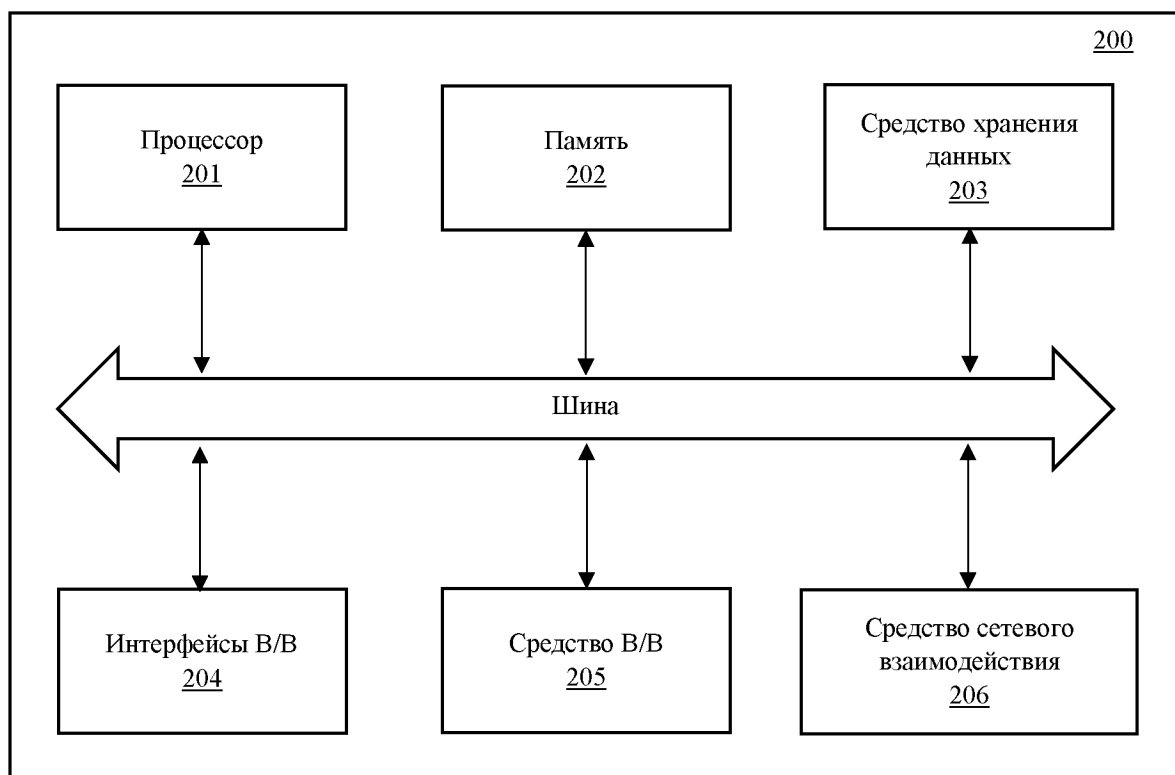
Новая программа лояльности даёт выгоду и удобство всем: и тем, кто с нами давно, и новичкам. Введение кэшбэка вместо скидки будет мотивировать пользователей возвращаться к нам. Кроме того, теперь переход на следующий уровень зависит от километража: то есть, сколько километров пользователь проехал через сервис Ситидрайв, столько и учитывается в его прогрессе. Это позволит сделать процесс перехода на новые уровни максимально прозрачным и понятным.

Алексей Сергеев

Генеральный директор Ситидрайв

Все подробности о новой программе лояльности доступны на официальном сайте Ситидрайв в разделе «Программа лояльности».

Фиг. 7



Фиг. 8

ОТЧЕТ О ПАТЕНТНОМ ПОИСКЕ
(статья 15(3) ЕАПК и правило 42 Патентной инструкции к ЕАПК)

Номер евразийской заявки:

202293405

А. КЛАССИФИКАЦИЯ ПРЕДМЕТА ИЗОБРЕТЕНИЯ:

G06F 16/951 (2019.01)

G06F 17/00 (2019.01)

G06F 40/30 (2020.01)

Согласно Международной патентной классификации (МПК)

Б. ОБЛАСТЬ ПОИСКА:

Просмотренная документация (система классификации и индексы МПК)

G06F 16/00, 16/951, 17/00, 40/00, 40/30

Электронная база данных, использовавшаяся при поиске (название базы и, если, возможно, используемые поисковые термины)
Google Patents, Espacenet, (ИС «Поисковая платформа» Роспатент), ЕАПАТИС

В. ДОКУМЕНТЫ, СЧИТАЮЩИЕСЯ РЕЛЕВАНТНЫМИ

Категория*	Ссылки на документы с указанием, где это возможно, релевантных частей	Относится к пункту №
D,A	US 2007/0198459 A1 (ACCENTURE GLOBAL SERVICES GMBH), 23.08.2007	1-5
A	EP 2599013 A1 (HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P), 05.06.2013	1-5
A	US 2016/0352772 A1 (CISCO TECHNOLOGY, INC), 01.12.2016	1-5
A	US 2015/0106157 A1 (ADOBE SYSTEMS INCORPORATED), 16.04.2015	1-5
A	US 2005/192936 A1 (MEEK CHRISTOPHER A et. al.), 01.09.2005	1-5

последующие документы указаны в продолжении

* Особые категории ссылочных документов:

«А» - документ, определяющий общий уровень техники

«D» - документ, приведенный в евразийской заявке

«E» - более ранний документ, но опубликованный на дату подачи евразийской заявки или после нее

«O» - документ, относящийся к устному раскрытию, экспонированию и т.д.

"P" - документ, опубликованный до даты подачи евразийской заявки, но после даты испрашиваемого приоритета"

«Т» - более поздний документ, опубликованный после даты приоритета и приведенный для понимания изобретения

«X» - документ, имеющий наиболее близкое отношение к предмету поиска, порочащий новизну или изобретательский уровень, взятый в отдельности

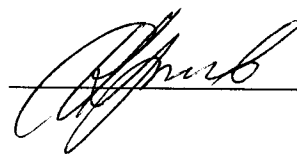
«Y» - документ, имеющий наиболее близкое отношение к предмету поиска, порочащий изобретательский уровень в сочетании с другими документами той же категории

«&» - документ, являющийся патентом-аналогом

«L» - документ, приведенный в других целях

Дата проведения патентного поиска: **13/04/2023**

Уполномоченное лицо:
Начальник отдела механики,
физики и электротехники



Д.Ф. Крылов