



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК

G06F 17/30598 (2006.01); *G06F 17/30997* (2006.01); *G06F 21/72* (2006.01)

(21)(22) Заявка: 2016114552, 29.12.2015

(24) Дата начала отсчета срока действия патента:
29.12.2015Дата регистрации:
16.03.2018

Приоритет(ы):

(30) Конвенционный приоритет:
17.11.2015 CN 201510791533.0

(43) Дата публикации заявки: 19.10.2017 Бюл. № 29

(45) Опубликовано: 16.03.2018 Бюл. № 8

(85) Дата начала рассмотрения заявки РСТ на
национальной фазе: 15.04.2016(86) Заявка РСТ:
CN 2015/099400 (29.12.2015)(87) Публикация заявки РСТ:
WO 2017/084179 (26.05.2017)Адрес для переписки:
191036, Санкт-Петербург, а/я 24, "НЕВИНПАТ"

(72) Автор(ы):

**ВАН Пинцзэ (CN),
ЧЗАН Тао (CN),
ЧЭНЬ Чжицзюнь (CN)**(73) Патентообладатель(и):
Сяоми Инк. (CN)(56) Список документов, цитированных в отчете
о поиске: US 6748367 B1, 08.06.2004. US
7908656 B1, 15.03.2011. WO 2012070931 A1,
31.05.2012. RU 2558617 C2, 10.08.2015.

(54) Способ и устройство для извлечения шаблона данных

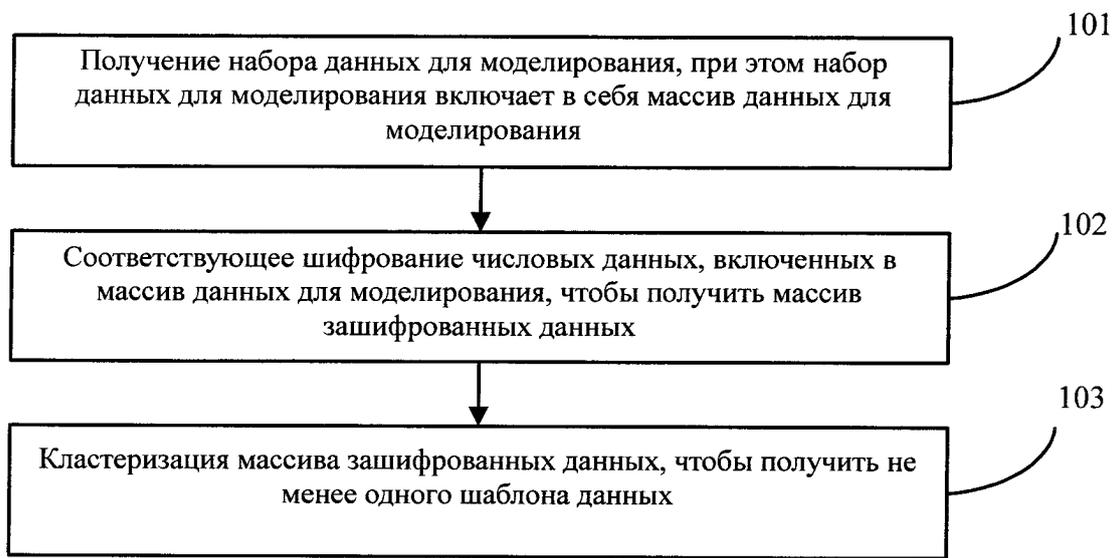
(57) Реферат:

Изобретение относится к извлечению шаблона данных. Техническим результатом является повышение точности обработки данных. Способ извлечения шаблона данных, включающий в себя: получение набора данных для моделирования, при этом набор данных для моделирования состоит из массива данных для моделирования; соответствующее шифрование числовых данных, входящих в массив данных для моделирования,

с целью получения массива зашифрованных данных, кластеризацию массива зашифрованных данных с целью получения не менее одного шаблона данных; подтверждение коэффициента охвата для каждого шаблона данных; и определение фактического шаблона данных на основании не менее одного шаблона данных в соответствии с результатами подтверждения. 3 н. и 20 з.п. ф-лы, 4 ил.

RU 2 647 628 C 2

RU 2 647 628 C 2



Фиг. 1



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(12) **ABSTRACT OF INVENTION**

(52) CPC

G06F 17/30598 (2006.01); *G06F 17/30997* (2006.01); *G06F 21/72* (2006.01)(21)(22) Application: **2016114552, 29.12.2015**(24) Effective date for property rights:
29.12.2015Registration date:
16.03.2018

Priority:

(30) Convention priority:
17.11.2015 CN 201510791533.0(43) Application published: **19.10.2017 Bull. № 29**(45) Date of publication: **16.03.2018 Bull. № 8**(85) Commencement of national phase: **15.04.2016**(86) PCT application:
CN 2015/099400 (29.12.2015)(87) PCT publication:
WO 2017/084179 (26.05.2017)Mail address:
191036, Sankt-Peterburg, a/ya 24, "NEVINPAT"

(72) Inventor(s):

**WANG Pingze (CN),
ZHANG Tao (CN),
CHEN Zhijun (CN)**

(73) Proprietor(s):

Xiaomi Inc. (CN)(54) **METHOD AND APPARATUS FOR MINING INFORMATION TEMPLATE**

(57) Abstract:

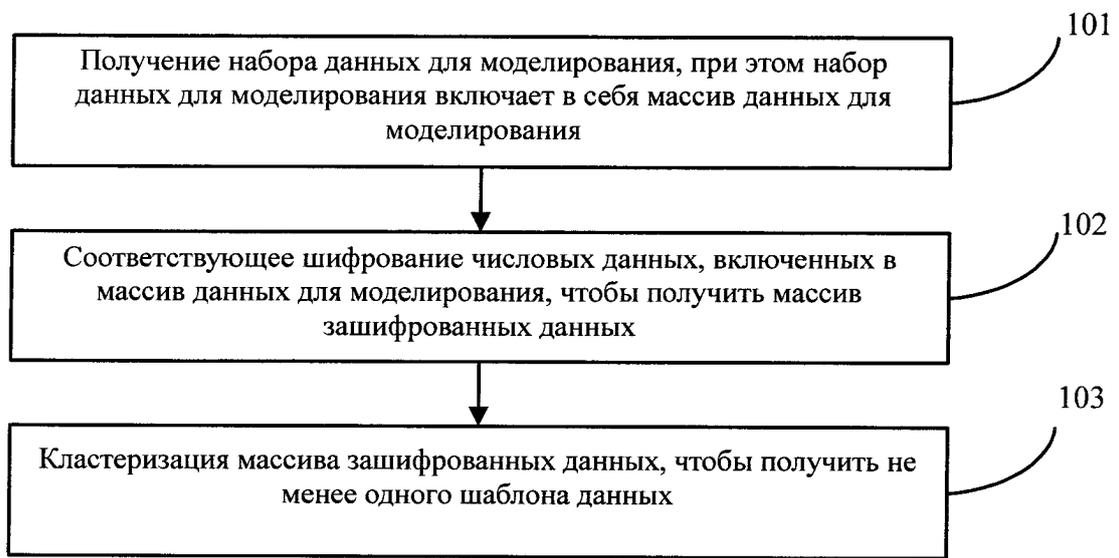
FIELD: information technology.

SUBSTANCE: invention relates to mining an information template. Method of mining an information template, including: obtaining a modeling information set, wherein the modeling information set comprises multiple pieces of modeling information; separately encrypting numerical information contained in the multiple pieces of modeling information to obtain multiple pieces of encrypted information, and clustering

the multiple pieces of encrypted information to obtain at least one information template; confirming the coverage factor for each information template; and determining the actual information template based on at least one information template in accordance with the confirmation results.

EFFECT: technical result is improved accuracy of information processing.

23 cl, 4 dwg



Фиг. 1

ПЕРЕКРЕСТНАЯ ССЫЛКА НА СВЯЗАННЫЕ ЗАЯВКИ

[0001] Настоящая заявка основана на Заявке №201510791533.0 на выдачу Китайского патента, все содержание которой включено в настоящий документ и по которой испрошен приоритет. Данная заявка была подана 17 ноября 2015 г.

ОБЛАСТЬ ТЕХНИКИ

[0002] Настоящее изобретение, в целом, относится к области информационных технологий, и, более конкретно, к способу и устройству для извлечения шаблона данных.

УРОВЕНЬ ТЕХНИКИ

[0003] В условиях современной жизни пользователи ежедневно получают большой объем уведомляющей информации, отправляемой в виде коротких сообщений, писем по электронной почте и мгновенных сообщений. В ходе исследования таких данных было определено, что, как правило, такие сообщения обладают схожими структурами текста, а шаблоны данных, извлеченные на основании данных структур текста, являются крайне важными для таких аспектов, как улучшение взаимодействия с пользователем, а также поддержка технологий обработки естественного языка.

[0004] В настоящее время извлечение часто встречающихся последовательностей представляет собой широко распространенный способ извлечения шаблона данных, в ходе которого осуществляют анализ массива данных на наличие шаблонов данных посредством предварительно заданных пороговых значений. Если при практическом применении частота обнаружения определенных данных превышает заданное пороговое значение, то в таком случае данные применяют в качестве шаблона данных.

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

[0005] Настоящим изобретением раскрыты способ и устройство для извлечения шаблона данных.

[0006] В соответствии с первым аспектом осуществления данного изобретения заявлен способ извлечения шаблона данных, включающий в себя:

[0007] получение набора данных для моделирования, при этом набор данных для моделирования включает в себя массив данных для моделирования;

[0008] соответствующее шифрование числовых данных, входящих в массив данных для моделирования, с целью получения массива зашифрованных данных, и

[0009] кластеризация массива зашифрованных данных с целью получения не менее одного шаблона данных.

[0010] В соответствии со вторым аспектом осуществления данного изобретения заявлено устройство извлечения шаблона данных, включающее в себя:

[0011] модуль сбора данных, сконфигурированный для получения набора данных для моделирования, при этом набор данных для моделирования включает в себя массив данных для моделирования;

[0012] модуль шифрования, сконфигурированный для соответствующего шифрования числовых данных, входящих в массив данных для моделирования, с целью получения массива зашифрованных данных, и

[0013] модуль кластеризации, сконфигурированный для кластеризации массива зашифрованных данных с целью получения не менее одного шаблона данных.

[0014] В соответствии с третьим аспектом осуществления данного изобретения заявлено устройство извлечения шаблона данных, включающее в себя:

[0015] процессор; и

[0016] запоминающее устройство для хранения команд, выполняемых процессором;

[0017] при этом процессор сконфигурирован для:

[0018] получения набора данных для моделирования, при этом набор данных для

моделирования включает в себя массив данных для моделирования;

[0019] соответствующего шифрования числовых данных, входящих в массив данных для моделирования, с целью получения массива зашифрованных данных, и

5 [0020] кластеризации массива зашифрованных данных с целью получения не менее одного шаблона данных.

[0021] Преимущества, обеспечиваемые техническими решениями в соответствии с вариантами осуществления данного изобретения, могут включать в себя:

10 [0022] шаблон данных извлекают путем анализа массива данных для моделирования, а шифрование числовых данных, входящих в данные для моделирования, происходит в ходе процесса извлечения шаблона, благодаря чему предотвращают утечку персональных данных пользователя вместе с извлеченным шаблоном. Таким образом обеспечивают более научный способ извлечения шаблона данных.

15 [0023] Следует понимать, что как вышеизложенное общее описание, так и последующее детальное описание являются всего лишь иллюстративными и поясняющими, и не ограничивают объем изобретения, как оно заявлено.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

[0024] Чертежи, которые включены в настоящую спецификацию и составляют ее часть, иллюстрируют варианты осуществления в соответствии с изобретением и, вместе с описанием, служат для разъяснения принципов изобретения.

20 [0025] Фиг. 1 представляет собой блок-схему извлечения шаблона данных в соответствии с примером осуществления.

[0026] Фиг. 2 представляет собой блок-схему извлечения шаблона данных в соответствии с примером осуществления.

25 [0027] Фиг. 3 представляет собой схематическое изображение устройства для извлечения шаблона данных в соответствии с примером осуществления.

[0028] Фиг. 4 представляет собой структурную блок-схему устройства для извлечения шаблона данных в соответствии с примером осуществления.

ПОДРОБНОЕ ОПИСАНИЕ

30 [0029] Далее приведено подробное описание примеров осуществления изобретения, проиллюстрированных на чертежах. Приведенное ниже описание относится к чертежам, на которых одни и те же номера на различных чертежах представляют одни и те же или аналогичные элементы, если на чертежах не указано иное. Варианты осуществления, представленные в последующем описании примеров осуществления, не представляют всех вариантов осуществления, соответствующих данному изобретению. Напротив, 35 они являются только примерами устройств и способов, соответствующих аспектам изобретения, как изложено в прилагаемых пунктах формулы изобретения.

[0030] Фиг. 1 представляет собой блок-схему извлечения шаблона данных в соответствии с примером осуществления. Как показано на Фиг. 1, способ извлечения шаблона данных применяют на сервере. Данный способ включает в себя следующие 40 шаги.

[0031] При выполнении шага 101 получают набор данных для моделирования, при этом набор данных для моделирования включает в себя массив данных для моделирования.

45 [0032] При выполнении шага 102 числовые данные, входящие в массив данных для моделирования, шифруют соответствующим образом с целью получения массива зашифрованных данных.

[0033] При выполнении шага 103 проводят кластеризацию массива зашифрованных данных с целью получения не менее одного шаблона данных.

[0034] В соответствии со способом, заявленным в данном варианте осуществления изобретения, извлечение (выделение) шаблона данных осуществляют путем анализа массива данных для моделирования, а шифрование числовых данных, входящих в данные для моделирования, происходит в ходе процесса извлечения шаблона, благодаря 5 чему предотвращают раскрытие персональных данных пользователя вместе с извлеченным шаблоном. Таким образом обеспечивают более научный способ для извлечения шаблона данных.

[0035] В другом варианте осуществления данного изобретения получение набора данных для моделирования включает в себя:

10 [0036] получение массива данных для обработки, при этом каждый блок данных, подлежащий обработке, по меньшей мере, содержит идентификатор передающей стороны;

[0037] разделение массива данных, подлежащего обработке, между различными наборами данных, при этом каждый блок данных, подлежащий обработке и включенный 15 в каждый из наборов данных, обладает одинаковым идентификатором передающей стороны;

[0038] получение первого предварительно заданного числа блоков данных, подлежащих обработке, в соответствии с идентификатором стороны, передающей 20 любой блок данных, подлежащий обработке, из соответствующего набора данных; и

[0039] составление набора данных для моделирования из первого предварительно заданного числа блоков данных, подлежащих обработке.

[0040] В другом варианте осуществления данного изобретения соответствующее шифрование числовых данных, включенных в массив данных для моделирования, с 25 целью получения массива зашифрованных данных включает в себя:

[0041] для любых данных для моделирования, - получение территориального идентификатора данных для моделирования, в соответствии с идентификатором 30 стороны, принимающей данные для моделирования;

[0042] шифрование идентификатора принимающей стороны, принимающей данных для моделирования с целью получения кода первого признака данных для 35 моделирования;

[0043] шифрование числовых данных из данных для моделирования с целью получения кода второго признака данных для моделирования, в зависимости от типа числовых данных, включенных в данные для моделирования; и

[0044] объединение в единый блок зашифрованных данных территориального 40 идентификатора, кода первого признака, кода второго признака и времени получения данных для моделирования.

[0045] В другом варианте осуществления данного изобретения кластеризация массива зашифрованных данных с целью получения не менее одного шаблона данных 45 выключает:

[0046] соответствующую консолидацию зашифрованных данных, которые удовлетворяют первому условию консолидации для массива зашифрованных данных, с целью получения первого массива консолидированных данных;

[0047] соответствующую консолидацию зашифрованных данных, которые удовлетворяют второму условию консолидации для первого массива 45 консолидированных данных, с целью получения второго массива консолидированных данных; и

[0048] кластеризацию второго массива консолидированных данных с целью получения не менее одного шаблона данных.

[0049] В другом варианте осуществления данного изобретения соответствующая консолидация зашифрованных данных, которые удовлетворяют первому условию консолидации для массива зашифрованных данных, с целью получения первого массива консолидированных данных включает в себя:

5 [0050] получение массива первых целевых блоков зашифрованных данных из массива зашифрованных данных, при этом зашифрованные данные в каждом первом целевом блоке зашифрованных данных имеют одинаковые территориальный идентификатор, код первого признака и код второго признака;

10 [0051] установку последовательности для зашифрованных данных в любом первом целевом блоке зашифрованных данных в зависимости от времени получения, в порядке от последнего до самого раннего; и

15 [0052] сохранение в первом целевом блоке зашифрованных данных, зашифрованных данных которые были получены последними, а также удаление остальных зашифрованных данных из первого целевого блока зашифрованных данных с целью получения единого первого консолидированного блока данных.

[0053] В другом варианте осуществления данного изобретения соответствующая консолидация зашифрованных данных, которые удовлетворяют второму условию консолидации для первого массива консолидированных данных с целью получения второго массива консолидированных данных включает в себя:

20 [0054] получение массива вторых целевых блоков зашифрованных данных из первого массива консолидированных данных, при этом первый блок консолидированных данных в каждом втором целевом блоке зашифрованных данных имеет одинаковые территориальный идентификатор и код второго признака;

25 [0055] для любого второго целевого блока зашифрованных данных, - получение числа первых блоков консолидированных данных во втором целевом блоке зашифрованных данных, а также последнего времени приема первых консолидированных данных во втором целевом блоке зашифрованных данных; и

30 [0056] объединение в единый второй блок консолидированных данных территориального идентификатора второго целевого блока зашифрованных данных, числа первых консолидированных данных во втором целевом блоке данных, кода второго признака второго целевого блока зашифрованных данных и последнего времени приема первых консолидированных данных во втором целевом блоке зашифрованных данных.

35 [0057] В другом варианте осуществления данного изобретения кластеризация второго массива консолидированных данных с целью получения не менее одного шаблона данных включает в себя:

40 [0058] разделение на массивы третьих целевых блоков зашифрованных данных второго массива консолидированных данных, при этом второй блок консолидированных данных в каждом третьем целевом блоке зашифрованных данных обладает одинаковым территориальным идентификатором;

[0059] для любого третьего целевого блока зашифрованных данных, - получение числа первых консолидированных данных, которые консолидированы по любым вторым консолидированным данным в третьем целевом блоке зашифрованных данных, и если число первых консолидированных данных, которые консолидированы по вторым консолидированным данным, превышает первое пороговое значение, объединение в единую запись информационного наполнения, соответствующего вторым консолидированным данным, числу первых консолидированных данных, которые консолидированы по вторым консолидированным данным, а также последнему времени

приема третьего целевого блока зашифрованных данных; и

5 [0060] консолидация массива записей, имеющих одинаковое информационное наполнение, из третьего целевого блока зашифрованных данных с целью получения одного шаблона данных. Число применимых зашифрованных данных в шаблоне данных представляет собой сумму числа зашифрованных данных, консолидированных для массива записей, а время приема шаблона данных представляет собой последнее время приема массива записей.

10 [0061] В другом варианте осуществления данного изобретения способ получения не менее одного шаблона данных после кластеризации массива зашифрованных данных далее включает в себя:

[0062] подтверждение коэффициента охвата каждого шаблона данных; и

[0063] определение фактического шаблона данных на основании не менее одного шаблона данных в соответствии с результатами подтверждения.

15 [0064] В другом варианте осуществления данного изобретения подтверждение коэффициента охвата каждого шаблона данных включает в себя:

[0065] получение набора тестовых данных, при этом набор тестовых данных включает в себя массив тестовых данных;

[0066] соответствующую обработку массива тестовых данных с целью получения первых блоков тестовых данных; и

20 [0067] подтверждение коэффициента охвата каждого шаблона данных в соответствии с массивом первых блоков тестовых данных.

[0068] В другом варианте осуществления данного изобретения получение набора тестовых данных включает в себя:

25 [0069] получение второго предварительно заданного числа подлежащих обработке данных с целью получения набора тестовых данных среди остальных подлежащих обработке данных из набора данных, соответствующих идентификатору стороны, отправляющей подлежащие обработке данные.

30 [0070] В другом варианте осуществления данного изобретения соответствующая обработка массива тестовых данных с целью получения первого массива блоков тестовых данных включает в себя:

[0071] соответствующее шифрование массива тестовых данных с целью получения массива зашифрованных тестовых данных;

35 [0072] соответствующую консолидацию зашифрованных тестовых данных, которые удовлетворяют третьему условию консолидации для массива зашифрованных тестовых данных, с целью получения третьего массива консолидированных данных; и

[0073] разделение третьего массива консолидированных данных на массивы первых целевых блоков тестовых данных, при этом третий блок консолидированных данных в каждом первом блоке тестовых данных обладает одинаковым территориальным идентификатором.

40 [0074] В другом варианте осуществления данного изобретения шифрование массива тестовых данных с целью получения массива зашифрованных тестовых данных включает в себя:

45 [0075] для любых тестовых данных в соответствии с идентификатором стороны, принимающей тестовые данные, - получение территориального идентификатора тестовых данных;

[0076] шифрование идентификатора стороны, принимающей тестовые данные с целью получения код первого признака тестовых данных; и

[0077] объединение в единый блок зашифрованных тестовых данных

территориального идентификатора тестовых данных, кода первого признака, и информационного наполнения тестовых данных.

[0078] В другом варианте осуществления данного изобретения консолидация массива зашифрованных тестовых данных с целью получения третьего массива консолидированных данных включает в себя:

[0079] получение второго массива блоков тестовых данных из массива зашифрованных тестовых данных, при этом зашифрованные тестовые данные в каждом втором блоке тестовых данных имеют одинаковые территориальный идентификатор и информационное наполнение;

[0080] получение числа зашифрованных тестовых данных в любом втором блоке тестовых данных; и

[0081] объединение в единый консолидированный блок тестовых данных территориального идентификатора второго блока тестовых данных, числа зашифрованных тестовых данных во втором блоке тестовых данных и информационного наполнения второго блока тестовых данных.

[0082] В другом варианте осуществления данного изобретения подтверждение коэффициента охвата каждого шаблона данных в соответствии с первым массивом блоков тестовых данных включает в себя:

[0083] с учетом любого шаблона данных - получение числа блоков данных, которые совпадают с шаблоном данных в каждом первом блоке тестовых данных; и

[0084] если число блоков данных в шаблоне данных, которые совпадают с любым первым блоком тестовых данных, превышает второе пороговое значение, - получение соотношения между числом совпавших блоков данных и числом блоков тестовых данных, включенных в первый блок тестовых данных.

[0085] В другом варианте осуществления данного изобретения определение фактического шаблона данных на основании не менее одного шаблона данных в соответствии с результатом подтверждения включает в себя:

[0086] если соотношение между числом совпавших данных и числом тестовых данных, включенных в первый блок тестовых данных, находится в пределах предварительно заданного диапазона, - определение того, что шаблон данных является фактическим шаблоном данных.

[0087] Все вышеуказанные дополнительные технические решения могут являться дополнительными вариантами осуществления данного изобретения при использовании в любом сочетании и не будут отдельно подробно раскрыты в настоящем документе.

[0088] Фиг. 2 представляет собой блок-схему способа извлечения шаблона данных в соответствии с примером осуществления. Как показано на Фиг. 2, способ извлечения шаблона данных применяют на сервере. Данный способ включает в себя следующие шаги.

[0089] При выполнении шага 201 сервер получает набор данных для моделирования, при этом набор данных для моделирования включает в себя массив данных для моделирования.

[0090] В условиях современной жизни пользователи ежедневно получают уведомляющую информацию, которую поставщики услуг отправляют в виде коротких сообщений, писем по электронной почте и мгновенных сообщений. При отправлении такой информации, как правило, используют единый шаблон данных, а сервера поставщиков услуг заполняют его соответствующими данными пользователей, после чего информация автоматически отправляется каждому из пользователей. Шаблоны данных, применяемые для таких данных, можно извлечь путем анализа таких данных.

Такие шаблоны данных могут не только помочь пользователям понять последние новости от поставщиков услуг и улучшить взаимодействие, но и могут способствовать развитию технологий обработки естественного языка.

5 [0091] Следует отметить, что поскольку различные типы данных относятся к разным поставщикам услуг, а информационное наполнение, отправляемое серверами различных поставщиков услуг, как правило, отличается, то чтобы провести целевое извлечение шаблона и повысить точность извлеченных шаблонов данных, представленный в варианте осуществления способ должен собирать данные одного типа в процессе извлечения шаблонов данных таким образом, чтобы извлечь шаблон данных из такого
10 типа данных.

[0092] Например, чтобы извлечь шаблон данных из собранных коротких сообщений, сервер может собирать различные короткие сообщения, отправляемые пользователям серверами таких поставщиков услуг, как «China Mobile», «China Telecom» и «China Unicom». Чтобы извлечь шаблон данных из собранных мгновенных сообщений, сервер
15 также может собирать мгновенные сообщения, отправляемые пользователям различными серверами приложений для обмена мгновенными сообщениями. Чтобы извлечь шаблон данных из собранных извещающих писем, отправленных по электронной почте, сервер также может собирать извещающие письма, отправляемые пользователям различными серверами электронной почты.

20 [0093] Объем собранных подлежащих обработке данных большой, и процесс извлечения шаблона данных будет крайне медленным, если извлечение шаблона данных будет осуществляться на основании всех собранных подлежащих обработке данных. Поэтому сервер может получить набор данных для моделирования на основании собранных подлежащих обработке данных. В набор данных для моделирования входит
25 совокупность данных для моделирования, при этом данные для моделирования представляют собой данные, подлежащие обработке с целью извлечения шаблона данных. Например, следующие шаги 2011-2014 могут быть применены, когда сервер получает набор данных для моделирования.

[0094] При выполнении шага 2011 сервер получает массив данных для обработки,
30 при этом каждый блок данных, подлежащий обработке, по меньшей мере, содержит идентификатор передающей стороны.

[0095] При сборе массива данных, подлежащих обработке, сервер может собирать данные, отправляемые пользователям сервером каждого из поставщиков услуг, чтобы
35 принять собранные данные в качестве полученного массива данных, которые подлежат обработке. В соответствии с вариантом осуществления каждый подлежащий обработке блок данных может включать, по меньшей мере, идентификатор передающей стороны, идентификатор принимающей стороны, данные о времени получения или схожую информацию. Если подлежащие обработке данные представляют собой короткие сообщения, то номер отправителя должен быть идентификатором передающей стороны,
40 который включен в каждый подлежащий обработке блок данных, а номер получателя должен быть идентификатором принимающей стороны. Если подлежащие обработке данные представляют собой сообщение электронной почты, то адрес почтового ящика, соответствующий серверу электронной почты, должен быть идентификатором передающей стороны, который включен в каждый подлежащий обработке блок данных,
45 а адрес почтового ящика пользователя должен быть идентификатором принимающей стороны. Если подлежащие обработке данные представляют собой мгновенное сообщение, то номер для обмена короткими сообщениями, соответствующий серверу приложений для обмена мгновенными сообщениями, должен быть идентификатором

передающей стороны, который включен в каждый подлежащий обработке блок данных, а номер для обмена короткими сообщениями пользователя или схожая информация должна быть идентификатором принимающей стороны.

5 [0096] Как правило, подлежащие обработке данные, собранные сервером, могли быть отправлены пользователю сервером каждого поставщика услуг в разное время, при этом некоторые из этих данных не имеют значения для пользователя, поскольку такие данные устарели, и по этой причине шаблоны данных, извлеченные на основании
10 таких данных, несомненно будут менее ценными. Таким образом, чтобы гарантировать актуальность извлеченных шаблонов данных, сервер может предварительно задать время после сбора массива подлежащих обработке данных, после чего в первую очередь
15 отфильтровывать подлежащие обработке данные в зависимости от заданного времени, чтобы сохранить подлежащие обработке данные со временем получения после заданного времени и удалять подлежащие обработке данные со временем получения до заданного времени.

15 [0097] При выполнении шага 2012 сервер распределяет подлежащий обработке массив данных между различными наборами данных, при этом каждый блок данных, подлежащий обработке и включенный в каждый из наборов данных, обладает одинаковым идентификатором передающей стороны.

[0098] В зависимости от идентификатора передающей стороны сервер разделяет
20 подлежащие обработке данные с одинаковым идентификатором передающей стороны в один набор данных, при этом подлежащие обработке данные в каждом наборе данных обладают одинаковым идентификатором передающей стороны. Благодаря разделению массива подлежащих обработке данных между каждым набором данных и идентификатором передающей стороны будет установлено взаимно-однозначное
25 соответствие. Например, подлежащие обработке данные с идентификатором передающей стороны 10086 могут быть распределены в набор данных А, а подлежащие обработке данные с идентификатором передающей стороны 10011 могут быть распределены в набор данных В или аналогичный набор данных.

[0099] При выполнении шага 2013 сервер получает первое предварительно заданное
30 число блоков подлежащих обработке данных из соответствующего набора данных в соответствии с идентификатором стороны, передающей любой блок подлежащих обработке данных.

[00100] В данном варианте осуществления между набором данных и идентификатором
35 передающей стороны установлено взаимно-однозначное соответствие, и сервер может получать соответствующий набор данных в зависимости от любого идентификатора передающей стороны. Поскольку каждый набор данных содержит большой объем подлежащих обработке данных, сложность операция высока, если извлечение шаблона данных будет осуществляться на основании всех подлежащих обработке данных. Таким образом, чтобы повысить скорость извлечения шаблона данных, в соответствии со
40 способом, заявленным в данном варианте осуществления, из полученного набора данных можно получать частичные подлежащие обработке данные и тем самым сформировать единый набор данных для моделирования. Если в качестве примера принять первое предварительно заданное число подлежащих обработке данных, то первым предварительно заданным числом может быть 1000, 2000, 3000 или иное число,
45 при этом в варианте осуществления не будет указано конкретное первое предварительно заданное число.

[00101] При выполнении шага 2014 сервер составляет набор данных для моделирования из первого предварительно заданного числа данных, подлежащих

обработке.

[00102] На основании полученного первого предварительно заданного числа подлежащих обработке данных сервер может составить набор данных для моделирования из первого предварительно заданного числа подлежащих обработке данных.

[00103] Следует отметить, что поскольку в варианте осуществления указаны различные подлежащие обработке данные, то для упрощения разграничения разных подлежащих обработке данных они будут распределены, в настоящем варианте осуществления, на данные для моделирования и тестовые данные в зависимости от различных способов применения подлежащих обработке данных. Данные для моделирования в основном применяют для извлечения шаблона данных, а тестовые данные в основном применяют для подтверждения извлеченных шаблонов данных.

[00104] При выполнении шага 202 сервер соответствующим образом шифрует числовые данные, входящие в массив данных для моделирования, с целью получения массива зашифрованных данных.

[00105] Обычно данные, получаемые пользователем, могут содержать некоторые персональные данные, например, имя, номер телефона, дату рождения, номер банковской карточки, номер членской карточки, суммы затрат, код подтверждения, номер чека об оплате или иные персональные данные. В настоящем варианте осуществления имя представляет собой текстовые персональные данные, а номер телефона, дата рождения, номер банковской карточки, номер членской карточки, суммы затрат, код подтверждения, номер чека об оплате или иные персональные данные принадлежат к числовым персональным данным, которые называют числовыми данными.

[00106] Текстовые персональные данные являются уникальными для пользователей, и различные текстовые персональные данные обычно отличаются или не полностью идентичны для различных пользователей. На основании текстовых данных не индивидуального характера, например, текста «уважаемые клиенты» или схожих текстов, которые часто содержатся в коротких сообщениях, невозможно определить идентификационную информацию пользователя, следовательно, такие данные не принадлежат к персональным данным. На основании текстовых данных индивидуального характера, например, Сань ЧЖАН и Сы ЛИ или аналогичного текста, можно получить соответствующие данные, например личность пользователя. Таким образом, чтобы предотвратить раскрытие персональных данных пользователя, сервер может выбрать подходящий символ подстановки для замены таких данных.

[00107] Чтобы улучшить защиту персональных данных пользователя, а именно числовых данных, включенных в данные для моделирования, сервер может зашифровать такие данные при помощи алгоритма шифрования для получения множества зашифрованных данных. Следующие шаги 2021-2024 могут быть применены для выполнения конкретного процесса.

[00108] При выполнении шага 2021 сервер получает территориальный идентификатор для данных для моделирования в соответствии с идентификатором стороны, принимающей данные для моделирования, для любых данных для моделирования.

[00109] Как правило, разные поставщики услуг используют разные шаблоны данных в разных местах или регионах с учетом различных особенностей регионов. Поскольку идентификатор принимающей стороны может определить местоположение или регионы различных пользователей, сервер может собирать территориальный идентификатор данных для моделирования в зависимости от идентификатора стороны, принимающей

данные для моделирования. В данном варианте осуществления территориальный идентификатор может быть представлен в качестве местоположения, и территориальные идентификаторы (местоположение), которые соответствуют различным местоположениям на территории, являются разными. Например, территориальным идентификатором Пекина является «пекин», а территориальным идентификатором Шанхая является «шанхай» или аналогичный идентификатор.

[00110] При выполнении шага 2022 сервер шифрует идентификатор стороны, принимающей данные для моделирования, чтобы получить код первого признака данных для моделирования.

[00111] При шифровании идентификатора стороны, принимающей данные для моделирования, сервер может применять предварительно заданный алгоритм шифрования, который определен сервером. Зашифрованный текст, соответствующий идентификатору стороны, принимающей данные для моделирования, может быть получен путем шифрования идентификатора стороны, принимающей данные для моделирования, при помощи предварительно заданного алгоритма шифрования. Зашифрованный текст называют кодом первого признака. В данном варианте осуществления код первого признака может быть представлен как шифротекст. Шифротекст отличается для разных идентификаторов принимающей стороны, например, шифротекстом идентификатора принимающей стороны 123456 является a, а шифротекстом идентификатора принимающей стороны 258369 является b или аналогичный шифротекст.

[00112] Следует отметить, что коды первого признака, полученные путем шифрования различных идентификаторов принимающей стороны при помощи предварительно заданного алгоритма шифрования, отличаются, а код первого признака, соответствующий каждому отдельному идентификатору принимающей стороны, является уникальным. Т.е. между идентификатором принимающей стороны и кодом первого признака установлено взаимно-однозначное соответствие.

[00113] При выполнении шага 2023 для получения кода второго признака сервер шифрует числовые данные из данных для моделирования в зависимости от типа числовых данных, включенных в данные для моделирования.

[00114] В данном варианте осуществления зашифрованные тексты, соответствующие разным типам числовых данных, также отличаются. Например, если типом числовых данных является номер телефона, то зашифрованным текстом, который соответствует числовым данным, может быть <Номер телефона>; если типом числовых данных является дата, то зашифрованным текстом, который соответствует числовым данным, может быть <Время>; а если вторые данные являются числовым типом, то зашифрованным текстом, который соответствует числовым данным, может быть <Номер>.

[00115] При шифровании числовых данных, включенных в данные для моделирования, сервер может применять зашифрованный текст, который соответствует типу числовых данных, чтобы заменить соответствующие числовые данные в данных для моделирования, и тем самым получить код второго признака данных для моделирования. В данном варианте осуществления код второго признака может быть представлен как наполнение Новое.

[00116] Например, если типом числовых данных из данных для моделирования является номер телефона 10086, то код второго признака <Номер телефона>, который соответствует номеру телефона, будет использован для замены 10086, чтобы получить код второго признака <Номер телефона> данных для моделирования.

[00117] При выполнении шага 2024 территориальный идентификатор, код первого признака, код второго признака и время получения данных для моделирования сервер составляет в единый блок зашифрованных данных.

5 [00118] Территориальный идентификатор, код первого признака, код второго признака и время получения данных для моделирования сервер формирует в единый блок зашифрованных данных. В данном варианте осуществления время приема может быть представлено как датавремя. В настоящее время зашифрованные данные могут быть четырехэлементными данными, которые можно представить как <местоположение, шифротекст, наполнениеНовое, датавремя>.

10 [00119] Следует отметить, что применение упомянутого выше шифрования любых данных для моделирования из набора данных для моделирования, например, в ходе процесса шифрования других данных для моделирования, является идентичным процессу шифрования данных для моделирования в ходе практического применения. Подробную информацию см. в вышеизложенном процессе шифрования данных для моделирования, 15 поскольку процесс не будет раскрываться далее в тексте.

[00120] Чтобы наглядно продемонстрировать процесс, который сервер выполняет с числовыми данными, включенными с массив данных для моделирования, далее приведено подробное пояснение и иллюстрации с коротким сообщением 1, коротким сообщением 2 и коротким сообщением 3 в качестве примеров.

20 [00121] Например, текст короткого сообщения 1 следующий: [Банк] Уважаемый(-ая) г-н/г-жа Фей ЛЮ, 21 августа 2014 г. Вы изменили свой предыдущий номер мобильного телефона (регистрационный номер 0102) на номер мобильного телефона (регистрационный номер 0988), просьба использовать новый номер мобильного телефона для осуществления транзакций посредством электронных банковских услуг, 25 а также звонить по номеру 95580 при возникновении любых вопросов.

[00122] Текст короткого сообщения 2 следующий: [Банк] Уважаемый(-ая) г-н/г-жа Цзи ЧЖАН, 13 июля 2014 г. Вы изменили свой предыдущий номер мобильного телефона (регистрационный номер 0209) на номер мобильного телефона (регистрационный номер 0898), просьба использовать новый номер мобильного телефона для осуществления транзакций посредством электронных банковских услуг, а также звонить по номеру 95580 при возникновении любых вопросов. 30

[00123] Текст короткого сообщения 3 следующий: [Банк] Уважаемый(-ая) г-н/г-жа Чанцзян ВАН, 07 августа 2014 г. Вы изменили свой предыдущий номер мобильного телефона (регистрационный номер 0109) на номер мобильного телефона (регистрационный номер 1234), просьба использовать новый номер мобильного телефона для осуществления транзакций посредством электронных банковских услуг, а также звонить по номеру 95580 при возникновении любых вопросов. 35

[00124] Поскольку в коротком сообщении 1, коротком сообщении 2 и коротком сообщении 3 содержатся такие персональные данные пользователя, как время, номер мобильного телефона, номер телефона или аналогичные персональные данные, то чтобы защитить персональные данные пользователя, можно использовать способ, изложенный в варианте осуществления, чтобы зашифровать числовые данные в таких коротких сообщениях. В результате шифрования могут быть получены следующие зашифрованные данные:

45 [00125] зашифрованные данные 1: [Банк] Уважаемый(-ая) г-н/г-жа Фей ЛЮ, <Время> Вы изменили свой предыдущий номер мобильного телефона (регистрационный номер <Номер>) на номер мобильного телефона (регистрационный номер <Номер>), просьба использовать новый номер мобильного телефона для осуществления транзакций

посредством электронных банковских услуг, а также звонить по номеру <Номер телефона> при возникновении любых вопросов;

5 [00126] зашифрованные данные 2: [Банк] Уважаемый(-ая) г-н/г-жа Цзи ЧЖАН, <Время> Вы изменили свой предыдущий номер мобильного телефона (регистрационный номер <Номер>) на номер мобильного телефона (регистрационный номер <Номер>), просьба использовать новый номер мобильного телефона для осуществления транзакций посредством электронных банковских услуг, а также звонить по номеру <Номер телефона > при возникновении любых вопросов; и

10 [00127] зашифрованные данные 3: [Банк] Уважаемый(-ая) г-н/г-жа Чанцзян ВАН, <Время> Вы изменили свой предыдущий номер мобильного телефона (регистрационный номер <Номер>) на номер мобильного телефона (регистрационный номер <Номер>), просьба использовать новый номер мобильного телефона для осуществления транзакций посредством электронных банковских услуг, а также звонить по номеру <Номер телефона> при возникновении любых вопросов.

15 [00128] Более того, чтобы упростить управление зашифрованными данными, сервер может составить набор зашифрованных данных из массива зашифрованных данных. В данном варианте осуществления набор зашифрованных данных может быть представлен как свод1. Свод1 включает в себя массив четырехэлементных зашифрованных данных, которые могут быть представлены как <местоположение, шифротекст, наполнениеНовое, датавремя>.

[00129] При выполнении шага 203 сервер выполняет кластеризацию массива зашифрованных данных с целью получения не менее одного шаблона данных.

25 [00130] Чтобы упростить процесс извлечения шаблонов данных и повысить эффективность извлечения шаблонов данных, сервер может на основании полученного массива зашифрованных данных выполнить кластеризацию массива зашифрованных данных и тем самым получить не менее одного шаблона данных.

[00131] В данном варианте осуществления процесс кластеризации сервером массива зашифрованных данных с целью получения не менее одного шаблона данных может быть продемонстрирован посредством следующих шагов 2031-2033.

30 [00132] При выполнении шага 2031 сервер для получения первого массива консолидированных данных выполняет соответствующую консолидацию зашифрованных данных, которые удовлетворяют первому условию консолидации для массива зашифрованных данных.

35 [00133] Первым условием консолидации является наличие одинаковых территориального идентификатора, кода первого признака и кода второго признака или иные параметры.

[00134] В качестве примера процесс сервера для соответствующей консолидации зашифрованных данных, которые удовлетворяют первому условию консолидации для массива зашифрованных данных, чтобы получить первый массив консолидированных данных, можно продемонстрировать посредством следующих шагов 20311-20313.

40 [00135] При выполнении шага 20331 сервер из массива зашифрованных данных получает массив первых целевых блоков зашифрованных данных, при этом зашифрованные данные в каждом первом целевом блоке зашифрованных данных имеют одинаковые территориальный идентификатор, код первого признака и код второго признака.

[00136] Поскольку все зашифрованные данные могут быть представлены в виде четырехэлементных данных, которые включают в себя территориальный идентификатор, код первого признака, код второго признака и время приема, сервер может получить

зашифрованные данные, которые обладают одинаковыми территориальным идентификатором, кодом первого признака и кодом второго признака в соответствии с формой четырехэлементного представления всех зашифрованных данных, чтобы сформировать первый целевой блок зашифрованных данных из блока зашифрованных данных, которые обладают одинаковыми территориальным идентификатором, кодом первого признака и кодом второго признака.

[00137] Например, сервер может получить массив зашифрованных данных, где зашифрованными данными 1 могут быть <пекин, а, Номер, 01.06.2015 г. 12:00:00>, зашифрованными данными 2 могут быть <пекин, а, Номер, 04.01.2015 г. 07:00:00>, зашифрованными данными 3 могут быть <пекин, а, Номер, 01.02.2015 г. 08:30>, зашифрованными данными 4 могут быть <тяньцзинь, с, НомерТелефона, 04.04.2015 г. 10:15:00>, зашифрованными данными 5 могут быть <тяньцзинь, с, НомерТелефона, 14.07.2015 г. 10:15:00>, и зашифрованными данными 6 могут быть <тяньцзинь, с, НомерТелефона, 12.07.2015 г. 06:20:00>. В данном варианте осуществления сервер формирует один первый целевой блок зашифрованных данных из зашифрованных данных 1, зашифрованных данных 2 и зашифрованных данных 3, которые имеют одинаковые территориальный идентификатор пекин, код первого идентификатора а и код второго идентификатора Номер. Сервер формирует другой первый целевой блок зашифрованных данных из зашифрованных данных 4, зашифрованных данных 5 и зашифрованных данных 6, которые имеют одинаковые территориальный идентификатор тяньцзинь, код первого идентификатора с и код второго идентификатора НомерТелефона.

[00138] При выполнении шага 20312 сервер сортирует зашифрованные данные в любом первом целевом блоке зашифрованных данных в зависимости от времени получения в нисходящем порядке (от последних до самых ранних).

[00139] Принимая в качестве примера зашифрованные данные в любом первом целевом блоке зашифрованных данных, сервер также может отсортировать зашифрованные данные в первом целевом блоке зашифрованных данных в зависимости от времени получения в нисходящем порядке.

[00140] Принимая в качестве примера первый целевой блок зашифрованных данных, который в вышеуказанном примере составлен из зашифрованных данных 1, зашифрованных данных 2 и зашифрованных данных 3, сервер сортирует зашифрованные данные в зависимости от времени получения в нисходящем порядке, чтобы получить зашифрованные данные 1, зашифрованные данные 2 и зашифрованные данные 3.

[00141] При выполнении шага 20313 сервер сохраняет зашифрованные данные в первом целевом блоке зашифрованных данных, который был получен последним, а также удаляет остальные зашифрованные данные из первого целевого блока зашифрованных данных, чтобы получить единый первый блок консолидированных данных.

[00142] Как правило, данные, обладающие одинаковыми территориальным идентификатором, кодом первого признака и кодом второго признака являются уведомляющей информацией, которая имеет одинаковое информационное наполнение и отправлена поставщиком услуг одному и тому же пользователю, при этом информация отправлена несколько раз по причине сбоя сервера. Чтобы при извлечении шаблонов данных снизить объем производимых вычислений для извлечения шаблонов данных, сервер может сохранить зашифрованные данные, которые были получены последними, в первом целевом блоке зашифрованных данных, и удалить остальные зашифрованные данные из первого целевого блока зашифрованных данных. Зашифрованные данные,

которые были получены последними, представляют собой первые консолидированные данные.

5 [00143] Например, если зашифрованными данными 1 в первом целевом блоке зашифрованных данных являются <пекин, а, Номер, 01.06.2015 г. 12:00:00>, зашифрованными данными 2 являются <пекин, а, Номер, 04.01.2015 г. 07:00:00>, зашифрованными данными 3 являются <пекин, а, Номер, 01.02.2015 г. 08:30> и зашифрованными данными 4 являются <тяньцзинь, с, НомерТелефона, 04.04.2015 г. 10:15:00>, то сервер может сохранить зашифрованные данные 1 и удалить зашифрованные данные 2 и зашифрованные данные 3.

10 [00144] При выполнении шага 2032 сервер для получения второго массива консолидированных данных выполняет соответствующую консолидацию зашифрованных данных, которые удовлетворяют второму условию консолидации для первого массива консолидированных данных.

15 [00145] В соответствии с одним из вариантов осуществления вторым условием консолидации является наличие одинаковых территории и кода второго признака или аналогичного параметра.

[00146] В качестве примера процесс, в котором сервер выполняет соответствующую консолидацию зашифрованных данных, которые удовлетворяют второму условию консолидации для первого массива консолидированных данных, чтобы получить второй массив консолидированных данных, можно продемонстрировать посредством 20 следующих шагов 20321-30323.

[00147] При выполнении шага 20321 из первого массива консолидированных данных сервер получает массив вторых целевых блоков зашифрованных данных, при этом 25 первый блок консолидированных данных в каждом втором целевом блоке зашифрованных данных имеет одинаковые территориальный идентификатор и код второго признака.

[00148] Поскольку каждые первые консолидированные данные могут быть 30 представлены в виде четырехэлементных данных, которые включают в себя территориальный идентификатор, код первого признака, код второго признака и время приема, сервер может получить первые консолидированные данные, которые обладают одинаковыми территориальным идентификатором и кодом второго признака в соответствии с формой четырехэлементного представления каждого зашифрованных данных, чтобы сформировать целевой блок зашифрованных данных из первых консолидированных данных, которые обладают одинаковыми территориальным 35 идентификатором и кодом второго признака.

[00149] Например, сервер соответствующим образом консолидирует зашифрованные данные, удовлетворяющие первому условию консолидации в массиве зашифрованных данных для получения первого массива зашифрованных данных, где первым условием консолидации 1 является <шанхай, с, Номер, 03.07.2015 г. 10:00:00>, первыми 40 консолидированными данными 2 являются <шанхай, а, Номер, 08.01.2015 г. 07:00:00>, первыми консолидированными данными 3 являются <пекин, б, НомерТелефона, 05.02.2015 г. 08:30:12>, первыми консолидированными данными 4 являются <шанхай, б, НомерТелефона, 18.07.2014 г. 09:00:02>, первыми консолидированными данными 5 являются <пекин, б, НомерТелефона, 04.05.2015 г. 10:20:00> и первыми 45 консолидированными данными 6 являются <тяньцзинь, с, НомерТелефона, 12.05.2015 г. 06:20:00>. Из первых консолидированных данных 1 и первых консолидированных данных 2, которые имеют одинаковые территориальный идентификатор шанхай, а также код второго признака Номер, сервер формирует единый второй целевой блок

зашифрованных данных. Из первых консолидированных данных 3 и первых консолидированных данных 5, которые имеют одинаковые территориальный идентификатор пекин и код второго признака НомерТелефона, сервер формирует единый второй целевой блок зашифрованных данных.

5 [00150] При выполнении шага 20322, с учетом любого второго целевого блока зашифрованных данных, сервер получает число первых блоков консолидированных данных во втором целевом блоке зашифрованных данных, а также последнее время приема первых консолидированных данных во втором целевом блоке зашифрованных данных.

10 [00151] В данном варианте осуществления число первых консолидированных данных во втором целевом блоке зашифрованных данных может быть равным 2, 3 или иному числу, а число первых консолидированных данных во вторых целевых блоках зашифрованных данных в данном варианте осуществления не задано. В данном варианте осуществления число первых консолидированных данных во втором целевом блоке
15 зашифрованных данных может быть представлено как с1.

[00152] В данном варианте осуществления сервер также сортирует первые консолидированные данные во второй целевой блок зашифрованных данных в зависимости от времени приема в порядке от последних до самых ранних, а также получает последнее время приема первых консолидированных данных во втором
20 целевом блоке зашифрованных данных в зависимости от результата сортировки. В данном варианте осуществления последнее время приема первых консолидированных данных во втором целевом блоке зашифрованных данных может быть представлено как последнеевремядата1.

[00153] Если в качестве примера принять вышеуказанный второй целевой блок
25 зашифрованных данных, составленный из первых консолидированных данных 3 и первых консолидированных данных 5, сервер может определить, что число первых консолидированных данных во втором целевом блоке зашифрованных данных равно 2. Поскольку временем получения первых консолидированных данных 3 является <05.02.2015 г. 08:30:12>, а временем получения первых консолидированных данных 5
30 является <04.05.2015 г. 10:20:00>, сервер может определить, что последним временем получения первых консолидированных данных во втором целевом блоке зашифрованных данных является 04.05.2015 г. 10:20:00.

[00154] При выполнении шага 20323 сервер из территориального идентификатора
35 второго целевого блока зашифрованных данных, числа первых консолидированных данных во втором целевом блоке зашифрованных данных, кода второго признака второго целевого блока зашифрованных данных и последнего времени приема вторых консолидированных данных во втором целевом блоке зашифрованных данных формирует единый второй блок консолидированных данных.

[00155] В соответствии с одним из вариантов осуществления каждые вторые
40 консолидированные данные являются четырехэлементными данными, которые могут быть представлены как <местоположение, с1, наполнениеНовое, последнеевремядата1>.

[00156] Следует отметить, что вышеизложенное объясняется консолидацией массива
45 первых консолидированных данных в едином втором целевом блоке зашифрованных данных. Процесс консолидации массива первых консолидированных данных в другом втором целевом блоке зашифрованных данных аналогичен вышеизложенному процессу консолидации массива первых консолидированных данных в любом втором целевом блоке зашифрованных данных и не будет рассмотрен в настоящем документе.

[00157] Чтобы упростить работу со вторым массивом консолидированных данных,

сервер также формирует единый набор свод2 из второго массива консолидированных данных.

[00158] При выполнении шага 2033 сервер выполняет кластеризацию второго массива консолидированных данных с целью получения не менее одного шаблона данных.

5 [00159] Несмотря на то, что посредством вышеизложенного процесса массив данных для моделирования в наборе данных для моделирования был консолидирован в виде второго массива консолидированных данных, число вторых консолидированных данных, полученных после консолидации, по-прежнему слишком велико, а частичные вторые консолидированные данные, включенные в эти вторые консолидированные
10 данные, по-прежнему обладают такими же признаками. Таким образом, сервер может выполнить кластеризацию второго массива консолидированных данных с целью получения не менее одного шаблона данных.

[00160] В другом варианте осуществления данного изобретения следующие шаги 20331-20333 могут быть применены, когда сервер выполняет кластеризацию второго
15 массива консолидированных данных, чтобы получить не менее одного шаблона данных.

[00161] При выполнении шага 20331 сервер разделяет второй массив консолидированных данных на массивы третьих целевых блоков зашифрованных данных, при этом второй блок консолидированных данных в каждом третьем целевом блоке зашифрованных данных обладает одинаковым территориальным
20 идентификатором.

[00162] Как правило, информационное наполнение, отправляемое поставщиками услуг пользователям, отличается в зависимости от разных признаков территории, и шаблоны данных, извлеченные сервером в зависимости от различного информационного наполнения, являются определенно разными. Таким образом, чтобы провести целевое
25 извлечение шаблонов данных, сервер может разделить второй массив консолидированных данных на третьи массивы целевых блоков зашифрованных данных в зависимости от территориального идентификатора вторых консолидированных данных, при этом вторые консолидированные данные в каждом третьем целевом блоке зашифрованных данных имеют одинаковый территориальный идентификатор.

30 [00163] В ходе конкретной реализации сервер может разделить набор свод2 на n отдельных подмножеств в зависимости от территориального идентификатора вторых консолидированных данных, включенных в свод2. Например, свод2 (местоположение 1), свод2 (местоположение1), свод2 (местоположение3), ..., свод2 (местоположениеn). С учетом любого свод2 (местоположениеi) территориальным идентификатором, который
35 соответствует каждому вторым консолидированным данным в свод2 (местоположение), является местоположениеi, а каждые вторые консолидированные данные в свод2 (местоположениеi) являются трехэлементными данными, которые могут быть представлены как <число, содержимоеНовое, датавремя>.

[00164] При выполнении шага 20332, для любого третьего целевого блока
40 зашифрованных данных сервер получает число зашифрованных данных, которые консолидированы по любым вторым консолидированным данным в третьем целевом блоке зашифрованных данных. И затем, если число зашифрованных данных, которые консолидированы по вторым консолидированным данным, превышает первое пороговое значение, сервер объединяет информационное наполнение, соответствующее вторым
45 консолидированным данным, число зашифрованных данных, которые консолидированы по вторым консолидированным данным, а также последнее время приема вторых зашифрованных данных в виде единой записи.

[00165] Чтобы гарантировать точность извлеченных шаблонов данных для любого

третьего целевого блока зашифрованных данных сервер может предварительно задать первое пороговое значение, которое может быть равным 10, 20, 30 или иному значению, чтобы на основании способа извлечения часто встречающихся подпоследовательностей извлечь шаблоны данных из целевого блока зашифрованных данных. При выполнении
5 такого процесса сервер может сравнивать число зашифрованных данных, консолидированных по любым вторым консолидированным данным в третьем целевом блоке зашифрованных данных, с пороговым значением для территории, и если число зашифрованных данных, консолидированных по вторым консолидированным данным, превышает первое пороговое значение, сервер формирует единую запись из
10 информационного наполнения, соответствующего вторым консолидированным данным, числа зашифрованных данных, консолидированных по вторым консолидированным данным, и последнего времени приема третьего целевого блока зашифрованных данных. Такая запись может быть обозначена как <посл, с2, последняяДатавремя2>, где посл
15 означает информационное наполнение, с2 означает число данных, консолидированных по третьим консолидированным данным, а последняяДатавремя2 означает последнее время приема вторых консолидированных данных.

[00166] При выполнении шага 20333 сервер объединяет массив записей, имеющих одинаковое информационное наполнение в третьем целевом блоке зашифрованных
20 данных, чтобы получить один шаблон данных. Число применимых зашифрованных данных в шаблоне данных представляет собой сумму числа зашифрованных данных, консолидированных для массива записей, а время приема шаблона данных представляет собой последнее время приема массива записей.

[00167] Сервер консолидирует массив записей, имеющих одинаковое информационное
25 наполнение из третьего целевого блока зашифрованных данных, для получения одного шаблона данных. Число применимых зашифрованных данных в шаблоне данных представляет собой сумму числа зашифрованных данных, консолидированных для массива записей, а время приема шаблона данных представляет собой последнее время приема массива записей.

[00168] Следует отметить, что вышеизложенное поясняется путем кластеризации
30 любого третьего целевого блока зашифрованных данных для получения одного шаблона данных. Вышеизложенный шаг 2033 применим для процесса кластеризации другого третьего целевого блока зашифрованных данных с целью получения шаблонов данных, и он не будет раскрыт в настоящем варианте осуществления.

[00169] На основании вышеизложенных шагов, с 201 по 203, можно понять, что
35 процесс извлечения шаблона данных, описанный в вариантах осуществления данного изобретения, в основном представляет собой процесс отделения усредненных шаблонов и сокрытия содержащихся в них таких персональных данных, как имена людей, наименование местоположения или иных текстовых данных, а также таких числовых
40 данных, как номер банковской карточки, номер мобильного телефона или иных числовых данных. Если в качестве примера принять извлечение шаблона данных для короткого сообщения 1, короткого сообщения 2 и короткого сообщения 3 в вышеизложенном шаге 2024, то применяя способ, описанный в вариантах осуществления данного изобретения, можно получить следующий шаблон данных:

[00170] [Банк] Уважаемый(-ая) г-н/г-жа <*>, <Время> Вы изменили свой предыдущий
45 номер мобильного телефона (регистрационный номер <Номер>) на номер мобильного телефона (регистрационный номер <Номер>), просьба использовать новый номер мобильного телефона для осуществления транзакций посредством электронных банковских услуг, а также звонить по номеру <НомерТелефона> при возникновении

любых вопросов.

[00171] При выполнении шага 204 сервер подтверждает коэффициент охвата каждого шаблона данных.

[00172] В целом, при выполнении вышеизложенного процесса извлечения шаблона данных можно полностью скрыть персональные данные пользователя в шаблоне. Однако, чтобы предотвратить неполное сокрытие персональных данных в шаблоне данных по причине программной ошибки, описанный в варианте осуществления метод будет также проводить подтверждение коэффициента покрытия для всех извлеченных шаблонов данных. Подтверждением коэффициента охвата называют вычисление коэффициента охвата шаблона данных в наборе тестовых данных, т.е. отношение числа совпавших данных относительно всех данных.

[00173] С учетом трех элементов <посл, с2, последнееВремядата2> каждого шаблона данных минимальный коэффициент охвата каждого шаблона данных в подлежащих обработке данных соответствующего свод2 (местоположениеi) теоретически является с. Если дано, что подлежащие обработке данные, и включенные в свод2 (местоположениеi) являются N, то коэффициент охвата посл в свод2 (местоположениеi) будет: коэффициент=c/N.

[00174] Как правило, ожидаемое значение коэффициента охвата для извлеченного шаблона данных посл в другом наборе данных также будет равным коэффициенту, при этом ожидаемое значение будет иметь отклонение ε. Т.е. ожидаемое значение коэффициента охвата для шаблона данных посл в другом наборе данных должно быть в диапазоне [коэффициент-ε, коэффициент+ε]. В случае посл с коэффициентом охвата в диапазоне [0, коэффициент-ε) его можно удалить из соответствующего шаблона данных, поскольку коэффициент охвата относительно мал. Дано пороговое значение охвата m, при этом посл со значением охвата менее m должно быть также удалено из соответствующего шаблона данных, чтобы коэффициент охвата фактического шаблона данных был в диапазоне [коэффициент-ε, 1] и имел достоверное абсолютное значение охвата.

[00175] Следует отметить, что процесс извлечения шаблона данных представляет собой процесс нахождения шаблона данных на основании подлежащих обработке данных, а процесс подтверждения коэффициента охвата представляет собой процесс применения шаблона для сопоставления с тестовыми данными. Эти два процесса являются взаимно противоположными, что может гарантировать отсутствие персональных данных пользователя в извлеченном шаблоне данных с учетом положительных и отрицательных аспектов. Более того, отсутствует перекрытие между набором данных, применяемым в процессе подтверждения коэффициента охвата, и набором данных, применяемым в процессе извлечения шаблона. Оба таких набора данных были случайным образом выделены из исходного набора данных. Две группы разных наборов данных, обладающих сходными распределениями, дополнительно гарантируют, что извлеченный шаблон данных не содержит персональных данных пользователя.

[00176] Например, следующие шаги 2041-2043 могут быть применены, когда сервер подтверждает коэффициент охвата каждого шаблона данных.

[00177] При выполнении шага 2041 сервер получает набор тестовых данных, при этом набор тестовых данных включает в себя массив тестовых данных.

[00178] Чтобы повысить точность результата подтверждения, сервер получит второе предварительно заданное число подлежащих обработке данных из остальных подлежащих обработке данных в наборе данных, который соответствует

идентификатору стороны, передающей подлежащие обработке данные, а затем сформирует набор тестовых данных из второго предварительно заданного числа подлежащих обработке данных. Вторым предварительно заданным числом может быть 100, 200 или 300 либо иное число, при этом второе предварительно заданное число

5 не указано точно в варианте осуществления.

[00179] При выполнении шага 2042 сервер соответствующим образом обрабатывает массив тестовых данных, чтобы получить первый массив блоков тестовых данных.

[00180] Следующие шаги 20421-20423 могут быть применены, когда сервер соответствующим образом обрабатывает массив тестовых данных, чтобы получить

10 первый массив блоков тестовых данных.

[00181] При выполнении шага 20421 сервер соответствующим образом шифрует массив тестовых данных, чтобы получить массив зашифрованных тестовых данных.

[00182] Следующие шаги 204211-204213 могут быть применены, когда сервер шифрует массив тестовых данных, чтобы получить массив зашифрованных тестовых данных.

15 [00183] При выполнении шага 204211 сервер для любых тестовых данных в соответствии с идентификатором стороны, принимающей тестовые данные, получает территориальный идентификатор для тестовых данных.

[00184] Как правило, разные поставщики услуг применяют разные шаблоны данных в разных местах или регионах с учетом различных особенностей регионов. Поскольку

20 идентификатор принимающей стороны может определять местоположения или регионы различных пользователей, сервер может получать идентификаторы территорий тестовых данных в зависимости от идентификатора стороны, принимающей тестовые данные.

В данном варианте осуществления территориальный идентификатор может быть представлен в качестве местоположения, а соответствующие идентификаторы

25 территорий (местоположение), которые отличаются друг от друга, если местоположения на территории являются разными. Например, территориальным идентификатором Пекина является «пекин», а территориальным идентификатором Шанхая является «шанхай» или иной идентификатор.

[00185] При выполнении шага 204212 сервер шифрует идентификатор стороны, принимающей тестовые данные, чтобы получить код первого признака тестовых

30 данных.

[00186] При шифровании идентификатора стороны, принимающей тестовые данные, сервер может применять предварительно заданный алгоритм шифрования, который определен сервером. Зашифрованный текст, соответствующий идентификатору стороны,

35 принимающей тестовые данные, может быть собран путем шифрования идентификатора стороны, принимающей тестовые данные, при помощи предварительно заданного алгоритма шифрования. Зашифрованный текст также называют кодом первого признака. В данном варианте осуществления код первого признака может быть

представлен как шифротекст. Шифротекст отличается для разных идентификаторов

40 принимающей стороны, например, шифротекстом идентификатора принимающей стороны 123456 является *a*, а шифротекстом идентификатора принимающей стороны 258369 является *s* или аналогичный шифротекст.

[00187] Следует отметить, что коды первого признака, полученные путем шифрования различных идентификаторов принимающей стороны при помощи предварительно

45 заданного алгоритма шифрования, отличаются, а код первого признака,

соответствующий каждому отдельному идентификатору принимающей стороны, является уникальным. Т.е. между идентификатором принимающей стороны и кодом первого признака установлено взаимно-однозначное соответствие.

[00188] При выполнении шага 204213 сервер составляет единый блок зашифрованных тестовых данных из территориального идентификатора тестовых данных, кода первого признака и информационного наполнения тестовых данных.

5 [00189] Сервер составляет единый блок зашифрованных данных из территориального идентификатора тестовых данных, кода первого признака и информационного наполнения тестовых данных. В данном варианте осуществления информационное наполнение тестовых данных может быть представлено как наполнение. В данный момент зашифрованные тестовые данные могут быть трехэлементными данными, которые можно представить как <местоположение, шифротекст, наполнение>.

10 [00190] Следует отметить, что применение упомянутого выше шифрования любых тестовых данных из набора тестовых данных, например, в ходе процесса шифрования других тестовых данных, является идентичным процессу шифрования тестовых данных в ходе практического применения. Подробную информацию см. в вышеизложенном процессе шифрования тестовых данных, поскольку процесс не будет раскрыт далее в

15 тексте.

[00191] При выполнении шага 20422 сервер для получения третьего массива консолидированных данных выполняет соответствующую консолидацию зашифрованных тестовых данных, которые удовлетворяют третьему условию консолидации для массива зашифрованных тестовых данных.

20 [00192] В ходе конкретной реализации могут быть применены следующие шаги 204221-204223 для того, чтобы сервер выполнял консолидацию массива зашифрованных тестовых данных с целью получения третьего массива консолидированных данных.

[00193] При выполнении шага 204221 сервер из массива зашифрованных тестовых данных получает второй массив блоков тестовых данных, при этом зашифрованные

25 тестовые данные в каждом втором блоке тестовых данных имеют одинаковый территориальный идентификатор и информационное наполнение.

[00194] Поскольку все зашифрованные тестовые данные могут быть представлены в виде трехэлементных данных, которые включают в себя территориальный идентификатор, код первого признака и информационное наполнение, сервер может

30 получить зашифрованные тестовые данные, которые обладают одинаковым территориальным идентификатором и информационным наполнением в соответствии с формой трехэлементного представления каждого зашифрованных тестовых данных, чтобы сформировать второй блок тестовых данных из блока зашифрованных данных, которые обладают одинаковым территориальным идентификатором и информационным

35 наполнением.

[00195] Например, сервер собирает массив зашифрованных тестовых данных, где зашифрованными тестовыми данными 1 являются <пекин, а, наполнение 1>, зашифрованными тестовыми данными 2 являются <пекин, b, наполнение 1>, зашифрованными тестовыми данными 3 являются <пекин, с, наполнение 1>, зашифрованными тестовыми данными 4 являются <тяньцзинь, d, наполнение 2> и зашифрованными тестовыми данными 5 являются <тяньцзинь, с, наполнение 2>. Сервер формирует один второй блок тестовых данных из зашифрованных тестовых данных 1, зашифрованных тестовых данных 2 и зашифрованных тестовых данных 3, которые имеют территориальный идентификатор пекин и информационное наполнение

45 наполнение 1. Сервер формирует другой второй блок тестовых данных из зашифрованных тестовых данных 4 и зашифрованных тестовых данных 5, которые имеют территориальный идентификатор тяньцзинь и информационное наполнение 2.

[00196] При выполнении шага 204222 сервер собирает число зашифрованных тестовых данных в любом втором блоке тестовых данных.

[00197] Число зашифрованных тестовых данных во втором блоке тестовых данных может быть равным 2, 3 или иному числу, а число зашифрованных тестовых данных во втором блоке тестовых данных в данном варианте осуществления не задано. В данном варианте осуществления число зашифрованных тестовых данных во втором блоке тестовых данных может быть представлено как с3.

[00198] При выполнении шага 204223 сервер формирует единый консолидированный блок тестовых данных из территориального идентификатора второго блока тестовых данных, числа зашифрованных тестовых данных во втором блоке тестовых данных и информационного наполнения второго блока тестовых данных.

[00199] Все вторые консолидированные данные являются трехэлементными данными, которые могут быть представлены как <местоположение, с3, наполнение>.

[00200] Следует отметить, что вышеизложенное объясняется объединением зашифрованных тестовых данных в любом втором блоке тестовых данных. Процесс объединения зашифрованных тестовых данных в другом втором блоке тестовых данных аналогичен вышеизложенному процессу объединения зашифрованных тестовых данных в любом втором блоке тестовых данных и не будет рассмотрен в настоящем документе.

[00201] Чтобы упростить работу с консолидированными тестовыми данными, сервер также формирует единый набор свод3 из массива консолидированных тестовых данных.

[00202] При выполнении шага 20423 сервер разделяет третий массив консолидированных данных на массивы первых целевых блоков тестовых данных, при этом третий блок консолидированных данных в каждом первом блоке тестовых данных обладает одинаковым территориальным идентификатором.

[00203] Как правило, информационное наполнение, отправляемое поставщиками услуг пользователям, отличается в зависимости от разных признаков территории, и шаблоны данных, которые могут быть подтверждены сервером в зависимости от различного информационного наполнения, являются определенно разными. Следовательно, чтобы провести целевое подтверждение шаблона данных, сервер может разделить третий массив консолидированных данных на массивы первых целевых блоков тестовых данных, при этом третий блок консолидированных данных в каждом первом блоке тестовых данных обладает одинаковым территориальным идентификатором. В ходе конкретной реализации сервер может разделить набор свод3 на n отдельных подмножеств в зависимости от территориального идентификатора вторых консолидированных данных, включенных в свод3. Например, подтвердить свод3 (местоположение 1), подтвердить свод3 (местоположение2), подтвердить свод3 (местоположение3), ... и подтвердить свод3 (местоположение n). В случае любого подтвердить свод3 (местоположение i) территориальным идентификатором, соответствующим каждому третьим консолидированным данным в подтвердить свод3 (местоположение i) является местоположение i , а каждые третьи консолидированные данные в подтвердить свод3 (местоположение i) являются трехэлементными данными, которые могут быть представлены как <местоположение, с4, наполнение>, при этом с4 представляет собой число третьих консолидированных данных, которые включены в каждый первый блок тестовых данных.

[00204] При выполнении шага 2043 сервер подтверждает коэффициент охвата каждого шаблона данных в соответствии с массивом первых блоков тестовых данных.

[00205] Следующие шаги 20431-20432 могут быть применены, когда сервер подтверждает коэффициент охвата для каждого шаблона данных в соответствии с

первым массивом блоков тестовых данных.

[00206] При выполнении шага 20431 для любого шаблона данных сервер собирает число блоков данных, которые совпадают с шаблоном данных в каждом первом блоке тестовых данных.

5 [00207] Если в качестве примера принять любой извлеченный шаблон данных, сервер также сопоставит шаблон данных с каждым первым блоком тестовых данных в первом массиве блоков тестовых данных. В ходе сопоставления наполнение шаблона данных может быть сопоставлено с информационным наполнением каждого третьих консолидированных данных, включенных в каждые первые тестовые данные. Если
10 наполнение шаблона данных идентично информационному наполнению любых третьих консолидированных данных, включенных в любые первые тестовые данные, или если схожесть между ними превышает конкретное значение, то считается, что шаблон данных совпадает с любыми третьими консолидированными данными, которые включены в
15 любые первые тестовые данные. Посредством такого способа сопоставления можно получить число данных в шаблоне данных, которые совпадают с каждым первым блоком тестовых данных в массиве первых блоков тестовых данных.

[00208] Чтобы более подробно пояснить вышеизложенный процесс, далее будет приведено подробное объяснение и иллюстрации с сопоставлением одного шаблона данных и одного первого блока тестовых данных в качестве примера.

20 [00209] Например, выделен следующий шаблон данных «[Банк] Уважаемый(-ая) г-н/г-жа <*>, <Время> Вы изменили свой предыдущий номер мобильного телефона (регистрационный номер <Номер>) на номер мобильного телефона (регистрационный номер <Номер>), просьба использовать новый номер мобильного телефона для осуществления транзакций посредством электронных банковских услуг, а также звонить
25 по номеру <НомерТелефона> при возникновении любых вопросов». Тестовыми данными, которые включены в первый блок тестовых данных, соответственно является короткое сообщение 1, короткое сообщение 2, короткое сообщение 3, короткое сообщение 4 и короткое сообщение 5, при этом коротким сообщением 1 является «
30 [Банк] Уважаемый(-ая) г-н/г-жа Фей ЛЮ, 21 августа 2014 г. Вы изменили свой предыдущий номер мобильного телефона (регистрационный номер 0102) на номер мобильного телефона (регистрационный номер 0988), просьба использовать новый номер мобильного телефона для осуществления транзакций посредством электронных банковских услуг, а также звонить по номеру 95580 при возникновении любых
35 вопросов»; коротким сообщением 2 является «[Банк] Уважаемый(-ая) г-н/г-жа Цзи ЧЖАН, 13 июля 2014 г. Вы изменили свой предыдущий номер мобильного телефона (регистрационный номер 0209) на номер мобильного телефона (регистрационный номер 8988), просьба использовать новый номер мобильного телефона для осуществления транзакций посредством электронных банковских услуг, а также звонить по номеру 95580 при возникновении любых вопросов»; коротким сообщением 3 является «[Банк]
40 Уважаемый(-ая) г-н/г-жа Чанцзян ВАН, 07 августа 2014 г. Вы изменили свой предыдущий номер мобильного телефона (регистрационный номер 0109) на номер мобильного телефона (регистрационный номер 1234), просьба использовать новый номер мобильного телефона для осуществления транзакций посредством электронных банковских услуг, а также звонить по номеру 95580 при возникновении любых
45 вопросов»; коротким сообщением 4 является «Уважаемый клиент, Вы можете избежать ввода ИД карточки или номера карточки, который нужно использовать в системе, позвонив по номеру мобильного телефона 186 (регистрационный номер 1234), чтобы предотвратить утомительный процесс управления и насладиться простотой

взаимодействия». Коротким сообщением 5 является «Уважаемый клиент, Ваш чек об оплате 10142389122 получен, и товар уже в пути, пожалуйста, ждите. Наш торговый центр не просит предоставлять данные вашей банковской карточки для возврата денег по причине недействительного чека об оплате. Опасайтесь мошенничества!» Если сопоставить шаблон данных с пятью короткими сообщениями в первом блоке тестовой информации можно определить, что короткое сообщение 1, короткое сообщение 2 и короткое сообщение 3 совпадают с шаблоном данных, а короткое сообщение 4 и короткое сообщение 5 не совпадают с шаблоном данных.

[00210] При выполнении шага 20432, если число данных, которые совпадают с любым первым блоком тестовых данных по шаблону данных, превышает второе пороговое значение, то сервер получает соотношение между числом совпавших данных и числом тестовых данных, включенных в первый блок тестовых данных.

[00211] Вторым пороговым значением может быть 3, 5, 8 или иное число, при этом второе пороговое значение не указано точно в варианте осуществления. Если число данных, совпадающих с любым первым блоком тестовых данных по шаблону данных, превышает второе пороговое значение, то сервер также получит соотношение между числом совпавших данных и числом тестовых данных, включенных в первый блок тестовых данных. Например, число данных, которые совпали с любым первым блоком тестовых данных по шаблону данных, равно t , а число тестовых данных, включенных в первый блок тестовых данных, равно o . В таком случае можно заключить, что соотношение между числом совпавших данных и числом тестовых данных, включенных в первый блок тестовых данных $=t/o$.

[00212] Например, если сопоставить шаблон данных с любым первым блоком тестовых данных сервер определяет, что число данных, совпавших с первым блоком тестовых данных по шаблону данных, равно 100. Если в качестве второго порогового значения задано 120, сервер может определить, что соотношение между числом совпавших данных и числом тестовых данных, включенных в первый блок тестовых данных, равно $100/120=83,3\%$.

[00213] При выполнении шага 205 сервер определяет фактический шаблон данных на основании не менее одного шаблона данных в соответствии с результатами подтверждения.

[00214] Когда способ, в соответствии с вышеизложенным шагом 204, применяют для подтверждения любого извлеченного шаблона данных, если соотношение между числом совпавших данных и числом тестовых данных, включенных в первый блок тестовых данных находится в пределах предварительно заданного диапазона, то такой шаблон данных можно считать фактическим шаблоном данных, где предварительно заданным диапазоном может быть [50%, 100%], [70%, 100%] или иной диапазон. Конкретный предварительно заданный шаблон не указан в настоящем варианте осуществления.

[00215] Более того, чтобы обеспечить 100% защиту персональных данных пользователя, после подтверждения коэффициента охвата шаблона данных сервер также предоставит подтвержденный фактический шаблон проверяющему персоналу для ручной проверки. Шаблон данных будет передан для применения соответствующему техническому персоналу только после успешной проверки шаблона.

[00216] В соответствии со способом, представленным в вариантах осуществления данного изобретения, шаблон данных извлекают путем анализа массива данных для моделирования, а шифрование числовых данных, входящих в данные для моделирования, происходит в ходе процесса извлечения шаблона, благодаря чему

предотвращают раскрытие персональных данных пользователя вместе с извлеченным шаблоном. Таким образом обеспечивают более научный способ для извлечения шаблона данных.

5 [00217] Фиг. 3 представляет собой схематическое изображение устройства для извлечения шаблона данных в соответствии с примером осуществления. В соответствии с Фиг. 3 устройство состоит из модуля 301 сбора данных, модуля 302 шифрования и модуля 303 кластеризации.

[00218] Модуль 301 сбора данных сконфигурирован для получения набора данных для моделирования, при этом набор данных для моделирования включает в себя массив
10 данных для моделирования.

[00219] Модуль 302 шифрования сконфигурирован для соответствующего шифрования числовых данных, входящих в массив данных для моделирования, с целью получения массива зашифрованных данных.

15 [00220] Модуль 303 кластеризации сконфигурирован для кластеризации массива зашифрованных данных с целью получения не менее одного шаблона данных.

[00221] В другом варианте осуществления данного изобретения модуль 301 сбора данных сконфигурирован для сбора массива подлежащих обработке данных, при этом подлежащие обработке данные включают в себя, по меньшей мере, идентификатор передающей стороны; последующего разделения массива подлежащих обработке
20 данных на разные наборы данных, при этом подлежащие обработке данные в каждом наборе данных имеют одинаковый идентификатор передающей стороны; последующего получения первого предварительно заданного числа подлежащих обработке данных из соответствующего набора данных в зависимости от идентификатора стороны, передающей любые подлежащие обработке данные; и составления набора данных для
25 моделирования из первого предварительно заданного числа подлежащих обработке данных.

[00222] В другом варианте осуществления данного изобретения модуль 302 шифрования сконфигурирован для любых данных для моделирования для сбора территориального идентификатора данных для моделирования в соответствии с
30 идентификатором стороны, принимающей данные для моделирования; последующего шифрования идентификатора стороны, принимающей данные для моделирования, чтобы получить код первого признака данных для моделирования; последующего получения кода второго признака данных для моделирования; шифрования числовых данных в данные для моделирования в зависимости от типа числовых данных, которые
35 включены в данные для моделирования; и объединения территориального идентификатора, кода первого признака, кода второго признака и времени приема данных для моделирования в единые зашифрованные данные.

[00223] В еще одном варианте осуществления данного изобретения модуль 303 кластеризации сконфигурирован для соответствующей консолидации зашифрованных
40 данных, которые удовлетворяют первому условию консолидации в массиве зашифрованных данных, чтобы получить первый массив консолидированных данных; выполнения соответствующей консолидации зашифрованных данных, которые удовлетворяют второму условию консолидации в первом массиве консолидированных данных, чтобы получить второй массив консолидированных данных; и выполнения
45 кластеризации второго массива консолидированных данных, чтобы получить не менее одного шаблона данных.

[00224] В еще одном варианте осуществления данного изобретения модуль 303 кластеризации сконфигурирован для сбора массива первых целевых блоков

зашифрованных данных из массива зашифрованных данных, при этом зашифрованные данные в каждом первом целевом блоке зашифрованных данных имеют одинаковые территориальный идентификатор, код первого признака и код второго признака; определения последовательности зашифрованных данных в любом первом целевом блоке зашифрованных данных в нисходящем порядке (т.е. от последних до самых ранних) в зависимости от времени приема; и сохранения зашифрованных данных в первом целевом блоке зашифрованных данных, который был получен последним, а также удаления остальных зашифрованных данных из первого целевого блока зашифрованных данных, чтобы получить единые первые консолидированные данные.

[00225] В еще одном варианте осуществления данного изобретения модуль 303 кластеризации сконфигурирован для сбора массива вторых целевых блоков зашифрованных данных из первого массива консолидированных данных, при этом первые консолидированные данные в каждом втором целевом блоке зашифрованных данных имеют одинаковые территориальный идентификатор и код второго признака; получения для любого второго целевого блока зашифрованных данных числа первых консолидированных данных во втором целевом блоке зашифрованных данных и последнего времени приема первых консолидированных данных во втором целевом блоке зашифрованных данных; и объединения территориального идентификатора второго целевого блока зашифрованных данных, числа первых консолидированных данных во втором целевом блоке консолидированных данных, кода второго признака второго целевого блока зашифрованных данных и последнего времени приема первых консолидированных данных во втором целевом блоке зашифрованных данных в единые вторые консолидированные данные.

[00226] В еще одном варианте осуществления данного изобретения модуль 303 кластеризации сконфигурирован для разделения второго массива консолидированных данных на массивы третьих целевых блоков зашифрованных данных, при этом вторые консолидированные данные в каждом третьем целевом блоке зашифрованных данных имеют одинаковый территориальный идентификатор; получения для любого третьего целевого блока зашифрованных данных числа первых консолидированных данных, которые консолидированы по любым вторым консолидированным данным в третьем целевом блоке зашифрованных данных; последующего объединения информационного наполнения, соответствующего вторым консолидированным данным, числа третьих целевых блоков зашифрованных данных, если число первых консолидированных данных, которые консолидированы по вторым консолидированным данным, превышает первое пороговое значение, в единую запись; и консолидации массива записей с одинаковым информационным наполнением в третьем целевом блоке зашифрованных данных, чтобы получить один шаблон данных. При этом число применимых зашифрованных данных шаблона данных представляет собой сумму числа зашифрованных данных, консолидированных по массиву записей, а время приема шаблона данных представляет собой последнее время приема массива записей.

[00227] В еще одном варианте осуществления данного изобретения устройство дополнительно включает в себя модуль подтверждения и модуль определения.

[00228] Модуль подтверждения сконфигурирован для подтверждения коэффициента охвата для каждого шаблона данных.

[00229] Модуль определения сконфигурирован для определения фактического шаблона данных на основании не менее одного шаблона данных в соответствии с результатами подтверждения.

[00230] В другом варианте осуществления данного изобретения модуль

подтверждения сконфигурирован для сбора набора тестовых данных, при этом набор тестовых данных включает в себя массив тестовых данных, соответствующей обработки массива тестовых данных для получения первого массива блоков тестовых данных и подтверждения коэффициента охвата каждого шаблона данных в соответствии с первым массивом блоком тестовых данных.

[00231] В другом варианте осуществления данного изобретения, чтобы получить набор тестовых данных, модуль подтверждения сконфигурирован для получения второго предварительно заданного числа подлежащих обработке данных из остальных подлежащих обработке данных в наборе данных, которые соответствуют идентификатору стороны, передающей подлежащие обработке данные.

[00232] В еще одном варианте осуществления данного изобретения модуль подтверждения сконфигурирован для соответствующего шифрования массива тестовых данных, чтобы получить массив зашифрованных тестовых данных; соответствующей консолидации зашифрованных тестовых данных, которые удовлетворяют третьему условию консолидации в массиве зашифрованных тестовых данных, чтобы получить третий массив консолидированных данных; и разделения третьего массива консолидированных данных между блоками тестовых данных первого массива, при этом третьи консолидированные данные в каждом первом блоке тестовых данных обладают одинаковым территориальным идентификатором.

[00233] В еще одном варианте осуществления данного изобретения модуль подтверждения сконфигурирован для любых тестовых данных для получения территориального идентификатора тестовых данных в зависимости от идентификатора стороны, принимающей тестовые данные; шифрования идентификатора стороны, принимающей тестовые данные, чтобы получить код первого признака тестовых данных; и объединения территориального идентификатора тестовых данных, кода первого признака и информационного наполнения тестовых данных в единые зашифрованные тестовые данные.

[00234] В еще одном варианте осуществления данного изобретения модуль подтверждения сконфигурирован для получения второго массива блоков тестовых данных из массива зашифрованных тестовых данных, при этом зашифрованные тестовые данные в каждом втором блоке тестовых данных имеют одинаковый территориальный идентификатор и информационное наполнение; получения числа зашифрованных тестовых данных в любом втором блоке тестовых данных и объединения территориального идентификатора второго блока тестовых данных, числа зашифрованных тестовых данных во втором блоке тестовых данных и информационного наполнения второго блока тестовых данных в единые консолидированные тестовые данные.

[00235] В еще одном варианте осуществления данного изобретения модуль подтверждения сконфигурирован для любого шаблона данных для получения числа данных, которые совпадают с шаблоном данных в каждом первом блоке тестовых данных и получения соотношения между числом совпавших данных и числом тестовых данных, которые включены в первый блок тестовых данных, если число данных, которые совпадают с шаблоном данных в каждом первом блоке тестовых данных, больше второго порогового значения.

[00236] В еще одном варианте осуществления данного изобретения модуль определения сконфигурирован для определения того, что шаблон данных является фактическим шаблоном данных, если соотношение между числом совпавших данных и числом тестовых данных, включенных в первый блок тестовых данных, находится в

пределах предварительно заданного диапазона.

[00237] В соответствии с устройством, представленным в вариантах осуществления данного изобретения, шаблон данных извлекают путем анализа массива данных для моделирования, а шифрование числовых данных, входящих в данные для моделирования, происходит в ходе процесса извлечения шаблона, благодаря чему предотвращают раскрытие персональных данных пользователя вместе с извлеченным шаблоном. Таким образом обеспечивают более научный способ для извлечения шаблона данных.

[00238] В отношении устройств, указанных в вышеизложенных вариантах осуществления, конкретные способы выполнения операций для отдельных модулей были подробно описаны в вариантах осуществления способов, которые не будут раскрыты в настоящем документе.

[00239] Фиг. 4 представляет собой блок-схему устройства 400 для извлечения шаблона данных в соответствии с примером осуществления. Например, устройством 400 может быть сервер. Как показано на Фиг. 4, устройство 400 включает в себя процессор 422, который дополнительно включает в себя один или несколько процессоров, а также ресурсы памяти, представляющие собой запоминающее устройство 432 для хранения команд, выполняемых процессором 422, например, прикладных программ. Прикладные программы, которые хранятся в запоминающем устройстве 432, могут включать в себя один или несколько модулей, каждый из которых соответствует набору команд. Процессор 422 дополнительно сконфигурирован на выполнение команд, чтобы осуществить вышеописанный метод для извлечения шаблона данных:

[00240] получение набора данных для моделирования, при этом набор данных для моделирования включает в себя массив данных для моделирования;

[00241] соответствующее шифрование числовых данных, входящих в массив данных для моделирования, с целью получения массива зашифрованных данных, а также

[00242] кластеризация массива зашифрованных данных с целью получения не менее одного шаблона данных.

[00243] В другом варианте осуществления изобретения получение набора данных для моделирования включает в себя:

[00244] получение массива данных для обработки, при этом каждый блок данных, подлежащий обработке, по меньшей мере, содержит идентификатор передающей стороны;

[00245] разделение массива данных, подлежащего обработке, между различными наборами данных, при этом каждый блок данных, подлежащий обработке и включенный в каждый из наборов данных, обладает одинаковым идентификатором передающей стороны;

[00246] получение первого предварительно заданного числа блоков данных, подлежащих обработке, в соответствии с идентификатором стороны, передающей любой блок данных, подлежащий обработке, из соответствующего набора данных; и

[00247] объединения первого предварительно заданного числа блоков данных, подлежащих обработке, в набор данных для моделирования.

[00248] В еще одном варианте осуществления данного изобретения соответствующее шифрование числовых данных, включенных в массив данных для моделирования, с целью получения массива зашифрованных данных включает в себя:

[00249] для любых данных для моделирования в соответствии с идентификатором стороны, принимающей данные для моделирования, - получение территориального идентификатора для данных для моделирования;

[00250] шифрование идентификатора стороны, принимающей данные для моделирования, с целью получения кода первого признака данных для моделирования;

[00251] шифрование числовых данных из данных для моделирования с целью получения кода второго признака данных для моделирования, в зависимости от числовых данных, включенных в данные для моделирования; и

[00252] объединения территориального идентификатора, кода первого признака, кода второго признака и времени получения данных для моделирования в единый блок зашифрованных данных.

[00253] В еще одном варианте осуществления изобретения кластеризация массива зашифрованных данных с целью получения не менее одного шаблона данных включает в себя:

[00254] соответствующую консолидацию зашифрованных данных, которые удовлетворяют первому условию консолидации для массива зашифрованных данных, с целью получения первого массива консолидированных данных;

[00255] соответствующую консолидацию зашифрованных данных, которые удовлетворяют второму условию консолидации для первого массива зашифрованных данных, с целью получения второго массива консолидированных данных; и

[00256] кластеризацию второго массива консолидированных данных с целью получения не менее одного шаблона данных.

[00257] В еще одном дополнительном варианте осуществления данного изобретения соответствующая консолидация зашифрованных данных, которые удовлетворяют первому условию консолидации для массива зашифрованных данных, с целью получения первого массива консолидированных данных включает в себя:

[00258] получение массива первых целевых блоков зашифрованных данных из массива зашифрованных данных, при этом зашифрованные данные в каждом первом целевом блоке зашифрованных данных имеют одинаковые территориальный идентификатор, код первого признака и код второго признака;

[00259] установку последовательности для зашифрованных данных в любом первом целевом блоке зашифрованных данных в зависимости от времени получения, в порядке от последнего до самого раннего; и

[00260] сохранение зашифрованных данных в первом целевом блоке зашифрованных данных, которые были получены последними, а также удаление остальных зашифрованных данных из первого целевого блока зашифрованных данных, с целью получения единого первого консолидированного блока данных.

[00261] В еще одном варианте осуществления данного изобретения соответствующая консолидация зашифрованных данных, которые удовлетворяют второму условию консолидации для первого массива консолидированных данных, с целью получения второго массива консолидированных данных включает в себя:

[00262] получение массива вторых целевых блоков зашифрованных данных из первого массива консолидированных данных, при этом первый блок консолидированных данных в каждом втором целевом блоке зашифрованных данных имеет одинаковые территориальный идентификатор и код второго признака;

[00263] для любого второго целевого блока зашифрованных данных - получение числа первых блоков консолидированных данных во втором целевом блоке зашифрованных данных, а также последнего времени приема первых консолидированных данных во втором целевом блоке зашифрованных данных; и

[00264] объединение территориального идентификатора из второго целевого блока зашифрованных данных, числа первых консолидированных данных во втором целевом

блоке данных, кода второго признака второго целевого блока зашифрованных данных и последнего времени приема первых консолидированных данных во втором целевом блоке зашифрованных данных в единый второй блок консолидированных данных.

5 [00265] В еще одном варианте осуществления данного изобретения кластеризация второго массива консолидированных данных с целью получения не менее одного шаблона данных включает в себя:

[00266] разделение второго массива консолидированных данных на массивы третьих целевых блоков зашифрованных данных, при этом второй блок консолидированных данных в каждом третьем целевом блоке зашифрованных данных обладает одинаковым
10 территориальным идентификатором;

[00267] для любого третьего целевого блока зашифрованных данных - получение числа первых консолидированных данных по любым вторым консолидированным данным в третьем целевом блоке зашифрованных данных и объединение
15 информационного содержимого по вторым консолидированным данным, числа первых консолидированных данных, которые консолидированы по вторым консолидированным данным, и последнего времени приема третьего целевого блока зашифрованных данных в единую запись, если число первых консолидированных данных, которые консолидированы по вторым консолидированным данным, превышает первое пороговое значение; и

20 [00268] консолидацию массива записей, имеющих одинаковое информационное наполнение, из третьего целевого блока зашифрованных данных с целью получения одного шаблона данных, при этом число применимых зашифрованных данных в шаблоне данных представляет собой сумму числа зашифрованных данных, консолидированных для массива записей, а время приема шаблона данных представляет
25 собой последнее время приема массива записей.

[00269] В еще одном варианте осуществления данного изобретения способ получения не менее одного шаблона данных после кластеризации массива зашифрованных данных далее включает в себя:

[00270] подтверждение коэффициента охвата каждого шаблона данных; и
30 [00271] определение фактического шаблона данных на основании не менее одного шаблона данных в соответствии с результатами подтверждения.

[00272] В еще одном варианте осуществления данного изобретения подтверждение коэффициента охвата каждого шаблона данных включает в себя:

[00273] получение набора тестовых данных, при этом набор тестовых данных
35 включает в себя массив тестовых данных;

[00274] соответствующую обработку массива тестовых данных, с целью получения первых блоков тестовых данных; и

[00275] подтверждение коэффициента охвата каждого шаблона данных в соответствии с массивом первых блоков тестовых данных.

40 [00276] В еще одном варианте осуществления данного изобретения получение набора тестовых данных включает в себя:

[00277] получение второго предварительно заданного числа подлежащих обработке данных с целью получения набора тестовых данных, среди остальных подлежащих
45 обработке данных из набора данных, соответствующих идентификатору стороны, отправляющей подлежащие обработке данные.

[00278] В еще одном варианте осуществления данного изобретения соответствующая обработка массива тестовых данных с целью получения первого массива блоков тестовых данных включает в себя:

[00279] соответствующее шифрование массива тестовых данных с целью получения массива зашифрованных тестовых данных;

5 [00280] соответствующую консолидацию зашифрованных тестовых данных, которые удовлетворяют третьему условию консолидации для массива зашифрованных тестовых данных, с целью получения третьего массива консолидированных данных; и

[00281] разделение третьего массива консолидированных данных на массивы первых целевых блоков тестовых данных, при этом третий блок консолидированных данных в каждом первом блоке тестовых данных обладает одинаковым территориальным идентификатором;

10 [00282] В еще одном варианте осуществления данного изобретения шифрование массива тестовых данных с целью получения массива зашифрованных тестовых данных включает в себя:

15 [00283] для любых тестовых данных - получение территориального идентификатора для тестовых данных в соответствии с идентификатором стороны, принимающей тестовые данные;

[00284] шифрование идентификатора стороны, принимающей тестовые данные, с целью получения кода первого признака тестовых данных; и

20 [00285] объединение территориального идентификатора тестовых данных, кода первого признака, информационного наполнения тестовых данных в единый блок зашифрованных тестовых данных.

[00286] В еще одном варианте осуществления данного изобретения консолидация массива зашифрованных тестовых данных с целью получения третьего массива консолидированных данных включает в себя:

25 [00287] получение второго массива блоков тестовых данных из массива зашифрованных тестовых данных, при этом зашифрованные тестовые данные в каждом втором блоке тестовых данных имеют одинаковый территориальный идентификатор и информационное наполнение;

[00288] получение числа зашифрованных тестовых данных в любом втором блоке тестовых данных; и

30 [00289] объединение территориального идентификатора второго блока тестовых данных, числа зашифрованных тестовых данных во втором блоке тестовых данных и информационного наполнения второго блока тестовых данных в единый консолидированный блок тестовых данных.

35 [00290] В другом варианте осуществления данного изобретения подтверждение коэффициента охвата каждого шаблона данных в соответствии с первым массивом блоков тестовых данных включает в себя:

[00291] для любого шаблона данных - получение числа блоков данных, которые совпадают с шаблоном данных в каждом первом блоке тестовых данных; и

40 [00292] если число данных, которые совпадают с любым первым блоком тестовых данных по шаблону данных, превышает второе пороговое значение, - получение соотношения между числом совпавших данных и числом тестовых данных, включенных в первый блок тестовых данных.

45 [00293] В еще одном варианте осуществления данного изобретения определение фактического шаблона данных на основании не менее одного шаблона данных в соответствии с результатом подтверждения включает в себя:

[00294] если соотношение между числом совпавших данных и числом тестовых данных, включенных в первый блок тестовых данных, находится в пределах предварительно заданного диапазона, - определение того, что шаблон данных является

фактическим шаблоном данных.

[00295] Устройство 400 также может включать в себя блок 426 питания, который сконфигурирован для управления питанием устройства 400, проводной или беспроводной сетевой интерфейс (интерфейсы) 450, который сконфигурирован для подключения устройства 400 к сети, а также интерфейс 458 ввода/вывода (I/O). Устройство 400 может работать на основании операционной системы, которая хранится в запоминающем устройстве 432, например Windows Server™, Mac OS X™, Unix™, Linux™, FreeBSD™ или иной операционной системе.

[00296] В соответствии с устройством, представленным в вариантах осуществления данного изобретения, шаблон данных извлекают путем анализа массива данных для моделирования, а шифрование числовых данных, входящих в данные для моделирования, происходит в ходе процесса извлечения шаблона, благодаря чему предотвращают раскрытие персональных данных пользователя вместе с извлеченным шаблоном. Таким образом обеспечивают более научный способ для извлечения шаблона данных.

[00297] Другие варианты осуществления изобретения будут очевидны для специалистов в данной области техники из рассмотрения описания и работы изобретения, описанных в настоящем документе. Предполагается, что данная заявка охватывает любые изменения, варианты использования или адаптации изобретения с соблюдением его общих принципов, включая такие отклонения от данного описания, которые возникают в рамках известной или традиционной практики в данной области техники. Предполагается, что описание и примеры считаются только примерами, при этом истинный объем изобретения указан в нижеследующей формуле изобретения.

[00298] Данное изобретение, предпочтительно, не ограничивается точной конструкцией, которая описана выше и проиллюстрирована на чертежах, и что различные модификации и изменения могут быть сделаны, оставаясь в пределах его объема. Предполагается, что объем изобретения ограничивается только прилагаемой формулой изобретения.

(57) Формула изобретения

1. Способ извлечения шаблона данных, включающий в себя:
 - получение набора данных для моделирования, при этом набор данных для моделирования состоит из массива данных для моделирования;
 - соответствующее шифрование числовых данных, входящих в массив данных для моделирования, с целью получения массива зашифрованных данных,
 - кластеризацию массива зашифрованных данных с целью получения не менее одного шаблона данных;
 - подтверждение коэффициента охвата для каждого шаблона данных; и
 - определение фактического шаблона данных на основании не менее одного шаблона данных в соответствии с результатами подтверждения.
2. Способ по п. 1, отличающийся тем, что получение набора данных для моделирования включает в себя:
 - получение массива данных для обработки, при этом каждый блок данных, подлежащий обработке, по меньшей мере, содержит идентификатор передающей стороны;
 - разделение массива данных, подлежащего обработке, между различными наборами данных, при этом каждый блок данных, подлежащий обработке и включенный в каждый из наборов данных, обладает одинаковым идентификатором передающей стороны;

получение первого предварительно заданного числа блоков данных, подлежащих обработке, в соответствии с идентификатором стороны, передающей любой блок данных, подлежащий обработке, из соответствующего набора данных; и

5 объединение первого предварительно заданного числа блоков данных, подлежащих обработке, в набор данных для моделирования.

3. Способ по п. 1, отличающийся тем, что соответствующее шифрование числовых данных, включенных в массив данных для моделирования, с целью получения массива зашифрованных данных включает в себя:

10 для любых данных для моделирования - получение территориального идентификатора для данных для моделирования в соответствии с идентификатором стороны, принимающей данные для моделирования;

шифрование идентификатора стороны, принимающей данные для моделирования, с целью получения кода первого признака данных для моделирования;

15 шифрование числовых данных из данных для моделирования с целью получения кода второго признака данных для моделирования, в зависимости от типа числовых данных, включенных в данные для моделирования; и

объединение территориального идентификатора, кода первого признака, кода второго признака и времени получения данных для моделирования в единый блок зашифрованных данных.

20 4. Способ по п. 1, отличающийся тем, что кластеризация массива зашифрованных данных с целью получения не менее одного шаблона данных включает в себя:

соответствующую консолидацию зашифрованных данных, которые удовлетворяют первому условию консолидации для массива зашифрованных данных с целью получения первого массива консолидированных данных;

25 соответствующую консолидацию зашифрованных данных, которые удовлетворяют второму условию консолидации для первого массива зашифрованных данных, с целью получения второго массива консолидированных данных; и

кластеризацию второго массива консолидированных данных с целью получения не менее одного шаблона данных.

30 5. Способ по п. 4, отличающийся тем, что соответствующая консолидация зашифрованных данных, которые удовлетворяют первому условию консолидации для массива зашифрованных данных, с целью получения первого массива консолидированных данных включает в себя:

35 получение массива первых целевых блоков зашифрованных данных из массива зашифрованных данных, при этом зашифрованные данные в каждом первом целевом блоке зашифрованных данных имеют одинаковые территориальный идентификатор, код первого признака и код второго признака;

40 установление последовательности для зашифрованных данных в любом первом целевом блоке зашифрованных данных в зависимости от времени получения, в порядке от последнего до самого раннего; и

сохранение зашифрованных данных в первом целевом блоке зашифрованных данных, которые были получены последними, а также удаление остальных зашифрованных данных из первого целевого блока зашифрованных данных, с целью получения одного первого консолидированного блока данных.

45 6. Способ по п. 4, отличающийся тем, что соответствующая консолидация зашифрованных данных, которые удовлетворяют второму условию консолидации для первого массива зашифрованных данных, с целью получения второго массива консолидированных данных включает в себя:

получение массива вторых целевых блоков зашифрованных данных из первого массива консолидированных данных, при этом первый блок консолидированных данных в каждом втором целевом блоке зашифрованных данных имеет одинаковые территориальный идентификатор и код второго признака;

5 с учетом любого второго целевого блока зашифрованных данных - получение числа первых блоков консолидированных данных во втором целевом блоке зашифрованных данных, а также последнего времени приема первых консолидированных данных во втором целевом блоке зашифрованных данных; и

10 объединение территориального идентификатора из второго целевого блока зашифрованных данных, числа первых консолидированных данных во втором целевом блоке зашифрованных данных, кода второго признака второго целевого блока зашифрованных данных и последнего времени приема первых консолидированных данных во втором целевом блоке зашифрованных данных в единый второй блок консолидированных данных.

15 7. Способ по п. 4, отличающийся тем, что кластеризация второго массива консолидированных данных для получения не менее одного шаблона данных включает в себя:

разделение второго массива консолидированных данных на массивы третьих целевых блоков зашифрованных данных, при этом второй блок консолидированных данных в 20 каждом третьем целевом блоке зашифрованных данных обладает одинаковым территориальным идентификатором;

25 для любого третьего целевого блока зашифрованных данных - получение числа первых консолидированных данных, которые консолидированы по любым вторым консолидированным данным в третьем целевом блоке зашифрованных данных, и объединение информационного наполнения, соответствующего вторым 30 консолидированным данным, числа первых консолидированных данных, которые консолидированы по вторым консолидированным данным, а также последнего времени приема третьего целевого блока зашифрованных данных в виде единой записи, если число первых консолидированных данных, которые консолидированы по вторым 35 консолидированным данным, превышает первое пороговое значение; и

консолидацию массива записей, имеющих одинаковое информационное наполнение, из третьего целевого блока зашифрованных данных с целью получения одного шаблона данных, при этом число зашифрованных данных, применимых к шаблону данных, представляет собой сумму числа зашифрованных данных, консолидированных для 40 массива записей, а время приема шаблона данных представляет собой последнее время приема массива записей.

8. Способ по п. 2, отличающийся тем, что подтверждение коэффициента для каждого шаблона данных включает в себя:

40 получение набора тестовых данных, при этом набор тестовых данных включает в себя массив тестовых данных;

соответствующую обработку массива тестовых данных с целью получения первых блоков тестовых данных; и

подтверждение коэффициента охвата для каждого шаблона данных в соответствии с массивом первых блоков тестовых данных.

45 9. Способ по п. 8, отличающийся тем, что получение набора тестовых данных включает в себя:

получение второго предварительно заданного числа подлежащих обработке данных с целью получения набора тестовых данных, среди остальных подлежащих обработке

данных из набора данных, соответствующих идентификатору стороны, отправляющей подлежащие обработке данные.

10. Способ по п. 8, отличающийся тем, что соответствующая обработка массива тестовых данных с целью получения первого массива блоков тестовых данных включает

5 в себя:
шифрование массива тестовых данных с целью получения массива зашифрованных тестовых данных;

соответствующую консолидацию зашифрованных тестовых данных, которые удовлетворяют третьему условию консолидации для массива зашифрованных тестовых

10 данных с целью получения третьего массива консолидированных данных; и
разделение третьего массива консолидированных данных на массивы первых целевых блоков тестовых данных, при этом третий блок консолидированных данных в каждом первом блоке тестовых данных обладает одинаковым территориальным идентификатором.

15 11. Способ по п. 10, отличающийся тем, что шифрование массива тестовых данных с целью получения массива зашифрованных тестовых данных включает в себя:

для любых тестовых данных - получение территориального идентификатора для тестовых данных в соответствии с идентификатором стороны, принимающей тестовые данные;

20 шифрование идентификатора стороны, принимающей тестовые данные с целью получения кода первого признака тестовых данных; и

объединение территориального идентификатора тестовых данных, кода первого признака, информационного наполнения тестовых данных в единый блок зашифрованных тестовых данных.

25 12. Способ по п. 10, отличающийся тем, что консолидация массива зашифрованных тестовых данных с целью получения третьего массива консолидированных данных включает в себя:

получение второго массива блоков тестовых данных из массива зашифрованных тестовых данных, при этом зашифрованные тестовые данные в каждом втором блоке

30 тестовых данных имеют одинаковые территориальный идентификатор и информационное наполнение;

получение числа зашифрованных тестовых данных в любом втором блоке тестовых данных; и

35 объединение территориального идентификатора второго блока тестовых данных, числа зашифрованных тестовых данных во втором блоке тестовых данных и информационного наполнения второго блока тестовых данных в единый консолидированный блок тестовых данных.

40 13. Способ по п. 8, отличающийся тем, что подтверждение коэффициента охвата для каждого шаблона данных в соответствии с первым массивом блоков тестовых данных включает в себя:

для любого шаблона данных - получение числа блоков данных, которые совпадают с шаблоном данных в каждом первом блоке тестовых данных; и

если число данных, которые совпадают с любым первым блоком тестовых данных по шаблону данных, превышает второе пороговое значение, - получение соотношения

45 между числом совпавших данных и числом тестовых данных, включенных в первый блок тестовых данных.

14. Способ по п. 13, отличающийся тем, что определение фактического шаблона данных на основании не менее одного шаблона данных в соответствии с результатом

подтверждения, включает в себя:

если соотношение между числом совпавших данных и числом тестовых данных, включенных в первый блок тестовых данных, находится в пределах предварительно заданного диапазона, - определение того, что определенный шаблон данных является фактическим шаблоном данных.

15. Устройство извлечения шаблона данных, включающее в себя:

модуль сбора данных, сконфигурированный для получения набора данных для моделирования, при этом набор данных для моделирования включает в себя массив данных для моделирования;

модуль шифрования, сконфигурированный для соответствующего шифрования числовых данных, входящих в массив данных для моделирования, с целью получения массива зашифрованных данных;

модуль кластеризации, сконфигурированный для кластеризации массива зашифрованных данных с целью получения не менее одного шаблона данных;

модуль подтверждения, сконфигурированный для подтверждения коэффициента охвата для каждого шаблона данных; и

модуль определения, сконфигурированный для определения фактического шаблона данных на основании не менее одного шаблона данных в соответствии с результатами подтверждения.

16. Устройство по п. 15, отличающееся тем, что модуль сбора данных сконфигурирован для получения массива подлежащих обработке данных, при этом все подлежащие обработке данные включают в себя, по меньшей мере, идентификатор передающей стороны; разделения массива подлежащих обработке данных на разные наборы данных, при этом подлежащие обработке данные в каждом наборе данных имеют одинаковый идентификатор передающей стороны; получения первого предварительно заданного числа подлежащих обработке данных из соответствующего набора данных в зависимости от идентификатора стороны, передающей любые подлежащие обработке данные, и объединения первого предварительно заданного числа подлежащих обработке данных в набор данных для моделирования.

17. Устройство по п. 15, отличающееся тем, что модуль шифрования сконфигурирован для любых данных для моделирования для получения территориального идентификатора данных для моделирования в соответствии с идентификатором стороны, принимающей данные для моделирования; шифрования идентификатора стороны, принимающей данные для моделирования, чтобы получить код первого признака данных для моделирования; шифрования числовых данных в данные для моделирования в зависимости от типа числовых данных, которые включены в данные для моделирования, чтобы получить код второго признака данных для моделирования; и объединения территориального идентификатора, кода первого признака, кода второго признака и времени приема данных для моделирования в единые зашифрованные данные.

18. Устройство по п. 15, отличающееся тем, что модуль кластеризации сконфигурирован для соответствующей консолидации зашифрованных данных, которые удовлетворяют первому условию консолидации в массиве зашифрованных данных, чтобы получить первый массив консолидированных данных; соответствующей консолидации зашифрованных данных, которые удовлетворяют второму условию консолидации в первом массиве консолидированных данных, чтобы получить второй массив консолидированных данных; и кластеризации второго массива консолидированных данных, чтобы получить не менее одного шаблона данных.

19. Устройство по п. 18, отличающееся тем, что модуль кластеризации также

сконфигурирован для получения массива первых целевых блоков зашифрованных данных из массива зашифрованных данных, при этом зашифрованные данные в каждом первом целевом блоке зашифрованных данных имеют одинаковый территориальный идентификатор, код первого признака и код второго признака; установления последовательности зашифрованных данных в любом первом целевом блоке зашифрованных данных в порядке от последних до самых ранних в зависимости от времени приема; сохранения зашифрованных данных в первом целевом блоке зашифрованных данных, который был получен последним, и удаление остальных зашифрованных данных из первого целевого блока зашифрованных данных, чтобы получить единые первые консолидированные данные.

20. Устройство по п. 18, отличающееся тем, что модуль кластеризации также сконфигурирован для получения массива вторых целевых блоков зашифрованных данных из первого массива консолидированных данных, при этом первые консолидированные данные в каждом втором целевом блоке зашифрованных данных имеют одинаковый территориальный идентификатор и код второго признака; для любого второго целевого блока зашифрованных данных, получения числа первых консолидированных данных во втором целевом блоке зашифрованных данных и последнего времени приема первых консолидированных данных во втором целевом блоке зашифрованных данных; и объединения территориального идентификатора второго целевого блока зашифрованных данных, числа первых консолидированных данных во втором целевом блоке консолидированных данных, кода второго признака второго целевого блока зашифрованных данных и последнего времени приема первых консолидированных данных во втором целевом блоке зашифрованных данных в единые вторые консолидированные данные.

21. Устройство по п. 18, отличающееся тем, что модуль кластеризации также сконфигурирован для распределения второго массива консолидированных данных на массивы третьих целевых блоков зашифрованных данных, при этом вторые консолидированные данные в каждом третьем целевом блоке зашифрованных данных имеют одинаковый территориальный идентификатор, для любого третьего целевого блока зашифрованных данных, получения числа первых консолидированных данных, которые консолидированы по любым вторым консолидированным данным в третьем целевом блоке зашифрованных данных, и объединения информационного наполнения, соответствующего вторым консолидированным данным, числа первых консолидированных данных, которые консолидированы по вторым консолидированным данным, и последнего времени приема третьего целевого блока зашифрованных данных в единую запись, если число первых консолидированных данных, которые консолидированы по вторым консолидированным данным, превышает первое пороговое значение, консолидации массива записей с одинаковым информационным наполнением в третьем целевом блоке зашифрованных данных, чтобы получить один шаблон данных, при этом число зашифрованных данных, применимых к шаблону данных, представляет собой сумму числа зашифрованных данных, консолидированных по массиву записей, а время приема шаблона данных представляет собой последнее время приема массива записей.

22. Устройство по п. 16, отличающееся тем, что модуль подтверждения также сконфигурирован для получения набора тестовых данных, при этом набор тестовых данных включает в себя массив тестовых данных; соответствующей обработки массива тестовых данных с целью получения первого массива блоков тестовых данных и подтверждения коэффициента охвата для каждого шаблона данных в соответствии с

первым массивом блоком тестовых данных.

23. Устройство извлечения шаблона данных, включающее в себя:

процессор; и

запоминающее устройство для хранения команд, выполняемых процессором;

5 при этом процессор сконфигурирован для:

получения набора данных для моделирования, при этом набор данных для моделирования состоит из массива данных для моделирования;

соответствующего шифрования числовых данных, входящих в массив данных для моделирования, с целью получения массива зашифрованных данных;

10 кластеризации массива зашифрованных данных с целью получения не менее одного шаблона данных;

подтверждения коэффициента охвата для каждого шаблона данных; и

определения фактического шаблона данных на основании не менее одного шаблона данных в соответствии с результатами подтверждения.

15

20

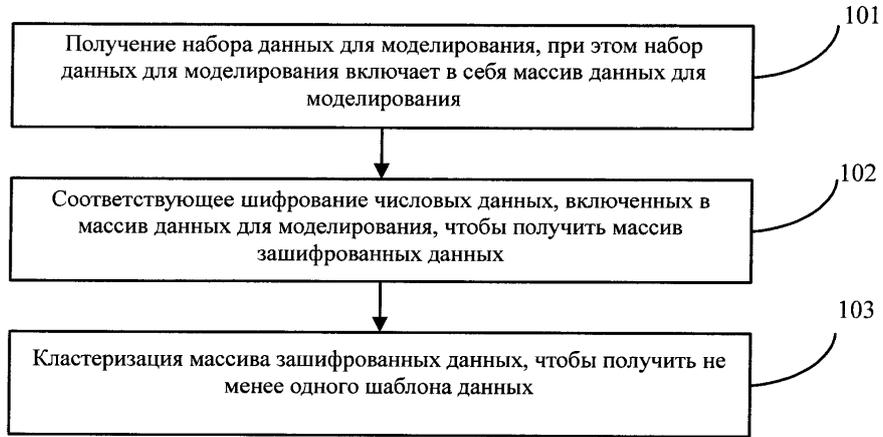
25

30

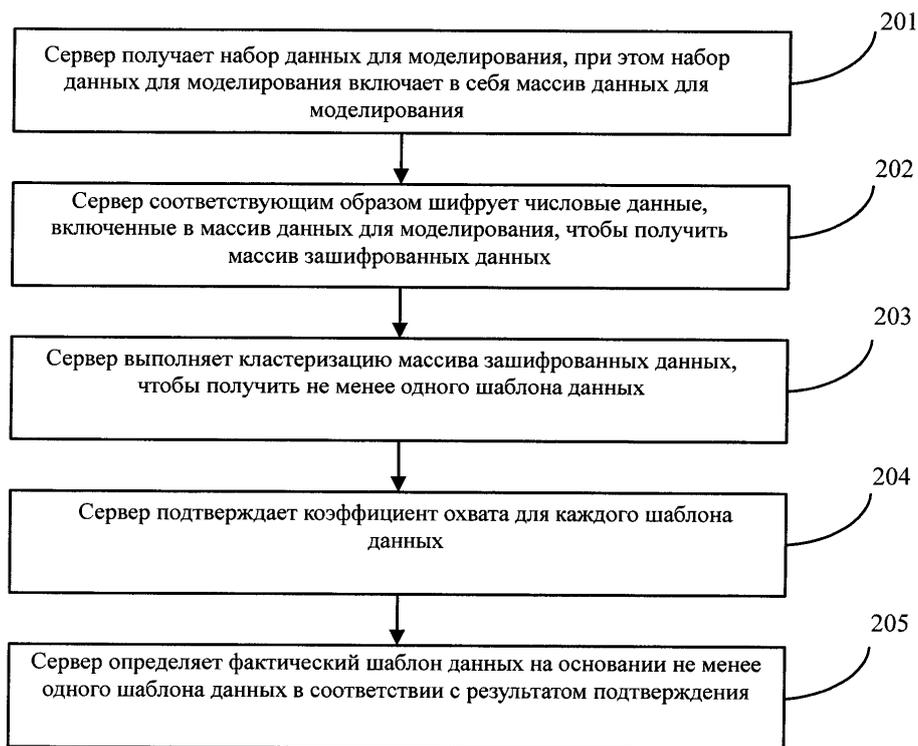
35

40

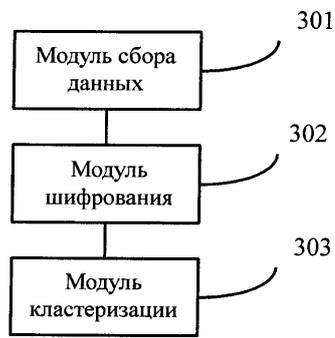
45



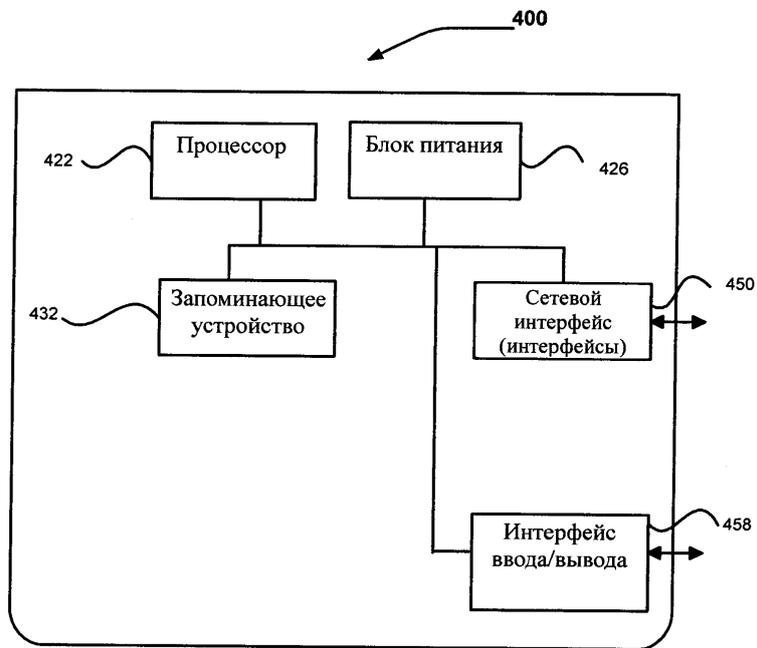
Фиг. 1



Фиг. 2



Фиг. 3



Фиг. 4