



(51) МПК  
*G10L 15/18* (2013.01)  
*G10L 15/30* (2013.01)  
*G06F 3/16* (2006.01)

ФЕДЕРАЛЬНАЯ СЛУЖБА  
 ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК

*G10L 15/18* (2019.02); *G10L 15/30* (2019.02); *G06F 3/16* (2019.02); *G10L 15/1822* (2019.02); *G10L 15/22* (2019.02); *G10L 15/265* (2019.02)

(21)(22) Заявка: 2018120514, 23.12.2016

(24) Дата начала отсчета срока действия патента:  
 23.12.2016

Дата регистрации:  
 31.10.2019

Приоритет(ы):

(30) Конвенционный приоритет:  
 01.03.2016 US 15/057,453

(45) Опубликовано: 31.10.2019 Бюл. № 31

(85) Дата начала рассмотрения заявки РСТ на  
 национальной фазе: 01.10.2018

(86) Заявка РСТ:  
 US 2016/068507 (23.12.2016)

(87) Публикация заявки РСТ:  
 WO 2017/151215 (08.09.2017)

Адрес для переписки:  
 129090, Москва, ул. Б.Спасская, 25, строение 3,  
 ООО "Юридическая фирма Городисский и  
 Партнеры"

(72) Автор(ы):

ВАН Бо (US),  
 ВЕМУРИ Сунил (US),  
 ДЖЕЙМС Барнаби Джон (US),  
 ХАФФМАН Скотт Б. (US),  
 ГУПТА Правир Кумар (US)

(73) Патентообладатель(и):  
 ГУГЛ ЭлЭлСи (US)

(56) Список документов, цитированных в отчете  
 о поиске: RU 2355044 C2, 10.05.2009. RU  
 2571520 C2, 20.12.2015. US 2011/0106534 A1,  
 05.05.2011. US 2015/0340042 A1, 26.11.2015. US  
 8849675 B1, 30.09.2014.

RU 2704876 C1

RU 2704876 C1

(54) СИСТЕМА ГОЛОСОВЫХ ДЕЙСТВИЙ РАЗРАБОТЧИКА

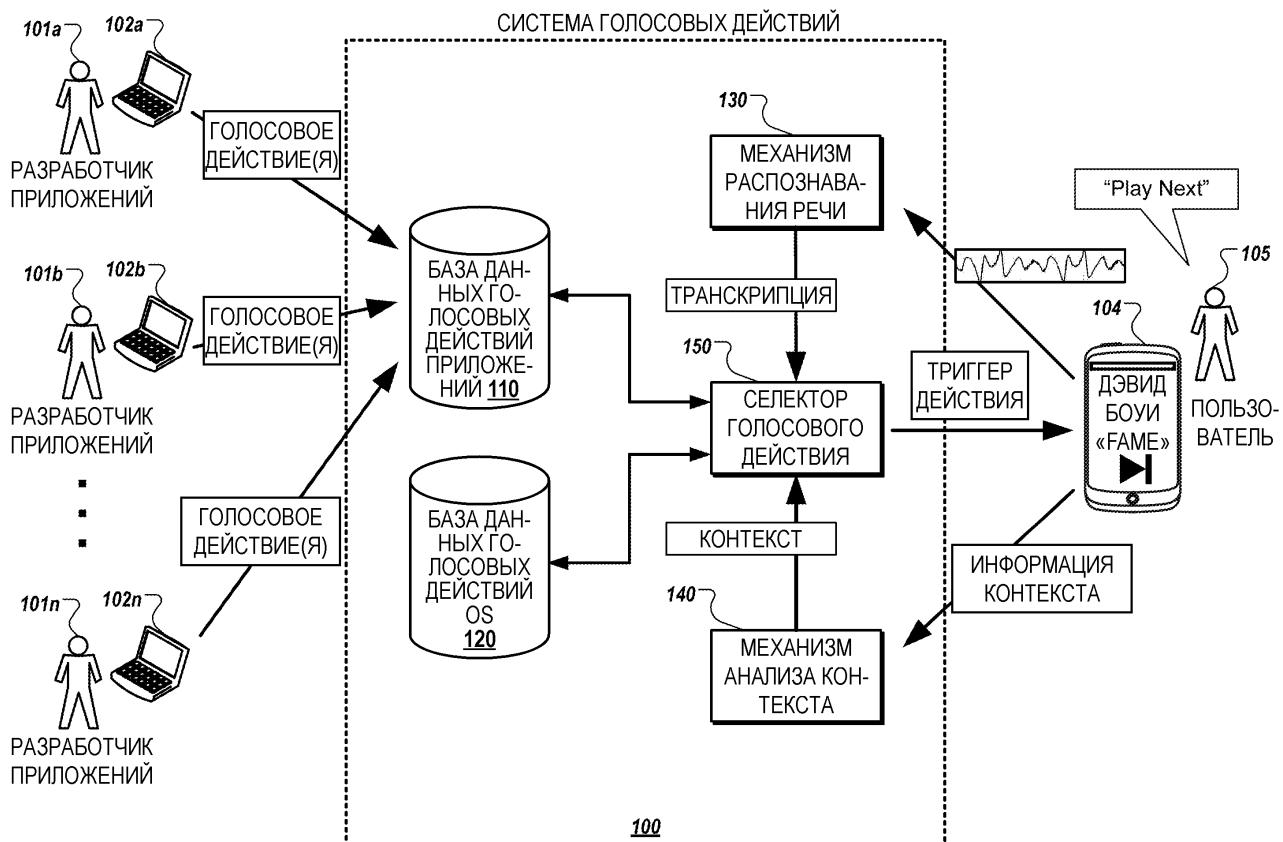
(57) Реферат:

Изобретение относится к области вычислительной техники. Технический результат заключается в обеспечении возможности развертывать новые голосовые действия для ранее установленных программных приложений. Технический результат достигается за счет приема от пользовательского устройства указания относительно высказывания пользователя, полученного пользовательским устройством, на котором установлено программное приложение, и контекстной информации, указывающей статус пользовательского устройства, на котором установлено программное приложение, или

программного приложения, установленного на пользовательском устройстве; определения, что транскрипция высказывания пользователя соответствует одному или более триггерным элементам, связанным с пассивной структурой данных голосового действия, и что контекстная информация удовлетворяет контексту для нового голосового действия; и в ответ на это определение подачи в пользовательское устройство пассивной структуры данных голосового действия, тем самым запрашивая, чтобы программное приложение, установленное на пользовательском устройстве, выполнило одну или более операций

нового голосового действия. З н. и 16 з.п. ф-лы,

4 ил.



ФИГ. 1

РУ 2704876 С 1



## (12) ABSTRACT OF INVENTION

(52) CPC

*G10L 15/18* (2019.02); *G10L 15/30* (2019.02); *G06F 3/16* (2019.02); *G10L 15/1822* (2019.02); *G10L 15/22* (2019.02); *G10L 15/265* (2019.02)

(21)(22) Application: 2018120514, 23.12.2016

(24) Effective date for property rights:  
23.12.2016Registration date:  
31.10.2019

Priority:

(30) Convention priority:  
01.03.2016 US 15/057,453

(45) Date of publication: 31.10.2019 Bull. № 31

(85) Commencement of national phase: 01.10.2018

(86) PCT application:  
US 2016/068507 (23.12.2016)(87) PCT publication:  
WO 2017/151215 (08.09.2017)

Mail address:

129090, Moskva, ul. B.Spasskaya, 25, stroenie 3,  
OOO "Yuridicheskaya firma Gorodisskij i  
Partnery"

(72) Inventor(s):

WANG Bo (US),  
VEMURI Sunil (US),  
JAMES Barnaby John (US),  
HUFFMAN Scott B. (US),  
GUPTA Pravir Kumar (US)

(73) Proprietor(s):

GOOGLE LLC (US)

## (54) DEVELOPER VOICE ACTIONS SYSTEM

(57) Abstract:

FIELD: computer equipment.

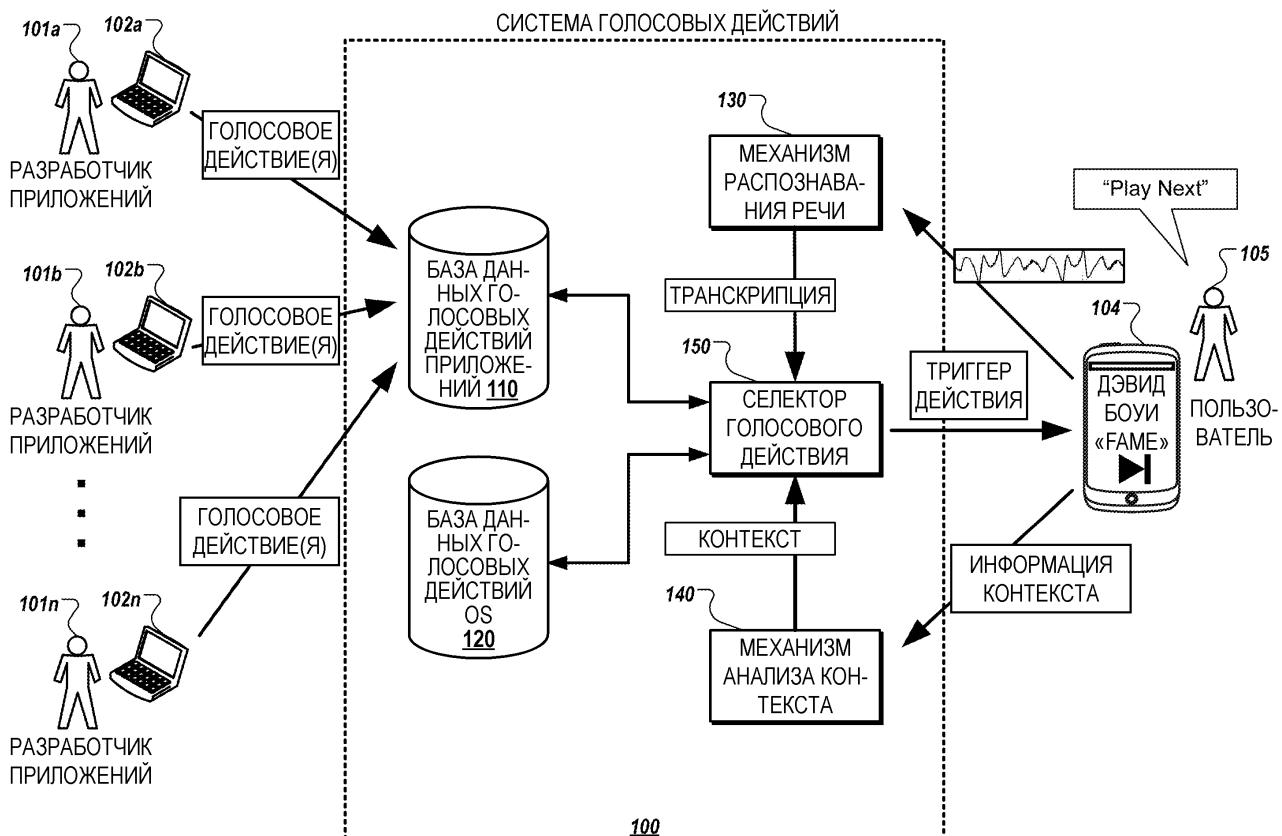
SUBSTANCE: invention relates to computer engineering. Technical result is achieved by receiving from a user device an indication of a user expression obtained by a user device, on which a software application is installed, and context information indicating the status of the user device on which the software application is installed, or a software application installed on the user device; determining that transcription of the user expression corresponds to one or more trigger elements associated with the passive

structure of the voice action data, and that the context information satisfies the context for the new voice action; and in response to this determination of the passive structure of voice action data to the user device, thereby requesting that the software application installed on the user device perform one or more new voice action operations.

EFFECT: possibility of deploying new voice actions for previously installed software applications.

19 cl, 4 dwg

RU 2704876 C1



## ФИГ. 1

## ПЕРЕКРЕСТНАЯ ССЫЛКА НА РОДСТВЕННУЮ ЗАЯВКУ

[0001] Эта заявка является продолжением и испрашивает приоритет заявки на патент США № 15/057,453, поданной 1 марта 2016, раскрытие которой полностью включено в настоящее описание путем ссылки.

### 5 ОБЛАСТЬ ТЕХНИКИ, К КОТОРОЙ ОТНОСИТСЯ ИЗОБРЕТЕНИЕ

[0002] Данное описание относится к голосовым действиям, и одна конкретная реализация относится к конфигурированию голосовых действий.

### УРОВЕНЬ ТЕХНИКИ

[0003] Задача в приложении может включать в себя одну или более активностей,

10 заданных в программном обеспечении, с которым пользователь взаимодействует для выполнения некоторой деятельности. Активность является классом, который управляет жизненным циклом исполнения задачи, так множество активностей, выполняемых в пределах задачи, могут позволить пользователю осуществлять деятельность. В некоторых реализациях команда может быть связана с активностью или действием, 15 относящимся к активности, так что представление пользователем команды может активировать активность или действие. Конкретное намерение может быть запущено, чтобы инициировать запуск активности или выполнения действия.

### СУЩНОСТЬ ИЗОБРЕТЕНИЯ

[0004] Данное описание описывает платформу, которая позволяет разработчику

20 приложений развертывать новые голосовые действия для ранее установленных программных приложений. Вообще, используя платформу, разработчик приложений может представить информацию, задающую новое голосовое действие, при этом информация указывает приложение, действие, которое новое голосовое действие должно активировать, триггерный элемент для активации действия и контекст, в котором 25 триггерный элемент должен активировать действия в приложении.

[0005] Контекст, представленный разработчиком приложений, может указывать статус пользовательского устройства, когда голосовое действие должно активировать действия в приложении. Статус устройства может включать в себя, например, какое 30 приложение(я) работает на переднем плане (то есть в настоящий момент активно в пользовательском интерфейсе устройства) или в фоновом режиме (то есть в настоящий момент скрыто из вида в пользовательском интерфейсе устройства), или может включать в себя конкретную информацию о приложениях, например, информацию о том, какую активность они в настоящий момент выполняют, статус выполняемой активности и так далее.

35 [0006] Когда разработчик приложений представляет информацию, задающую новое голосовое действие, служба или инструмент могут проверить информацию, задающую новое голосовое действие, чтобы определить, совместимо ли новое голосовое действие с приложением, или чтобы иным образом определить, может ли новое голосовое действие быть реализовано. Если новое голосовое действие допустимо, новый экземпляр 40 структуры пассивных данных, называемой намерением и имеющей конкретный формат, может быть создан для голосового действия путем ввода информации, задающей новое голосовое действие. Намерение может указывать некоторую или всю информацию, задающую новое голосовое действие, такую как приложение, триггерный элемент, активность или действие, которое должно быть активировано в ответ на обнаружение 45 триггерного элемента, и контекст, который необходим для активации триггерным элементом активности или действия. Затем голосовое действие может быть развернуто для приложения, так что голосовое действие становится разрешенным голосовым действием для приложения без выполнения дополнительных изменений в коде

программы для поддержки голосового действия.

[0007] Как только голосовое действие было развернуто, пользователь, эксплуатирующий пользовательское устройство, может обеспечить голосовой ввод. Пользовательское устройство может представить контекстную информацию для 5 пользователяского устройства, контекстная информация и транскрипция голосового ввода могут использоваться для идентификации намерения и для активации идентифицированного намерения. Когда намерение активировано, определяются данные для исполнения активности или действия и передаются пользовательскому устройству для исполнения активности или действия в ответ на голосовой ввод.

10 [0008] Например, разработчик приложения медиапроигрывателя может задать новое голосовое действие для перехода к следующей песне, которое использует триггерный элемент «play next». Разработчик приложений может указать контекст, когда голосовое действие «play next» должно быть разрешено. Например, разработчик приложений может указать, что голосовое действие «play next» должно вызвать переход приложения

15 медиапроигрывателя к следующей песне только тогда, когда приложение медиапроигрывателя выполняется на переднем плане и находится в режиме, который вызывает работу приложения медиапроигрывателя как аудиопроигрывателя. После того, как голосовое действие было развернуто разработчиком приложений, пользователь, имеющий ранее установленное приложение медиапроигрывателя на его

20 пользовательском устройстве, может представить голосовой ввод «play next». Пользовательское устройство может представить контекстную информацию, указывающую статус пользовательского устройства или приложений, установленных на пользовательском устройстве, например, контекстную информацию, указывающую, что приложение медиапроигрывателя выполняется на переднем плане и находится в

25 режиме аудиопроигрывателя. В ответ на прием голосового ввода «play next» и информации, указывающей контекст, данные могут быть переданы пользовательскому устройству, что вызывает переход приложения медиапроигрывателя к следующей песне. В противоположность этому, если определено, что в голосовом вводе сказано «play next», но контекстная информация указывает, что на переднем плане пользовательского

30 устройства выполняется приложение социальной сети вместо приложения медиапроигрывателя, то голосовой ввод «play next» может не иметь никакого эффекта или может вызывать выполнение другой операции в пользовательском устройстве.

[0009] Новаторские аспекты предмета изобретения, описанного в этом описании, могут быть воплощены в способах, которые включают в себя действия, содержащие 35 этапы, на которых принимают с помощью системы голосовых действий данные, указывающие новое голосовое действие для программного приложения, отличающегося от упомянутой системы голосовых действий, данные, содержащие одну или более операций для выполнения нового голосового действия и один или более триггерных элементов для активации нового голосового действия, генерируются с помощью системы

40 голосовых действий контекстное намерение голосового действия для программного приложения на основе по меньшей мере принятых данных, при этом контекстное намерение голосового действия содержит данные, которые при приеме программным приложением запрашивают, чтобы программное приложение выполнило одну или более операций нового голосового действия, связывают с помощью системы голосовых

45 действий контекстное намерение голосового действия с одним или более триггерными элементами для нового голосового действия, при этом система голосовых действий выполнена с возможностью: приема указания относительно высказывания пользователя, полученного устройством, на котором установлено программное приложение,

определения, что транскрипция высказывания пользователя соответствует одному или более триггерным элементам, связанным с контекстным намерением голосового действия, и, в ответ на это определение, предоставления устройству контекстного намерения голосового действия, тем самым запрашивая, чтобы программное

5 приложение, установленное на устройстве, выполнило одну или более операций нового голосового действия.

[0010] Каждый из этих и других вариантов осуществления optional могут включать в себя один или более из следующих признаков. В различных примерах принятые данные указывают контекст, этот контекст указывает статус устройства или

10 программного приложения, когда новое голосовое действие разрешено; контекст указывает, что программное приложение работает на переднем плане устройства, на котором установлено программное приложение; контекст указывает, что программное приложение работает в фоновом режиме устройства, на котором установлено

15 программное приложение; контекст указывает, что программное приложение выполняет конкретную активность; контекст указывает, что конкретная активность, которую выполняет программное приложение, находится в конкретном состоянии активности.

[0011] В других примерах каждый из этих и других вариантов осуществления optional могут включать в себя признаки, содержащие: связывание контекстного намерения голосового действия с контекстом для нового голосового действия, прием

20 системой голосовых действий контекстной информации, указывающей статус конкретного устройства, имеющего установленное программное приложение, или программного приложения, установленного на конкретном устройстве, определение, что контекстная информация удовлетворяет контексту для нового голосового действия, и, в ответ на определение, что транскрипция высказывания пользователя, полученная

25 конкретным устройством, соответствует одному или более триггерным элементам, связанным с контекстным намерением голосового действия, и что контекстная информация удовлетворяет контексту, связанному с контекстным намерением голосового действия, предоставление системой голосовых действий конкретному устройству контекстного намерения голосового действия, тем самым запрашивая,

30 чтобы программное приложение, установленное на конкретном устройстве, выполнило одну или более операций нового голосового действия; прием контекстной информации, указывающей статус конкретного устройства или программного приложения, установленного на конкретном устройстве, содержит: предоставление системой голосовых действий конкретному устройству запроса на конкретную контекстную

35 информацию, и прием конкретной контекстной информации в ответ на запрос.

[0012] В других примерах каждый из этих и других вариантов осуществления могут включать в себя один или более признаков, содержащих: определение, что контекстная информация удовлетворяет контексту для второго голосового действия, и что транскрипция высказывания пользователя соответствует одному или более триггерным

40 элементам, которые связаны с намерением для второго голосового действия, при этом намерение для второго голосового действия указывает одну или более операций для выполнения второго голосового действия, в ответ на это определение, выбор голосового действия между новым голосовым действием и вторым голосовым действием, и предоставление системой голосовых действий конкретному устройству намерения,

45 связанного с выбранным голосовым действием, тем самым запрашивая, чтобы программное приложение, установленное на конкретном устройстве, выполнило одну или более операций выбранного голосового действия; выбор выбранного голосового действия между новым голосовым действием и вторым голосовым действием содержит

выбор выбранного голосового действия в ответ на прием данных, указывающих выбор пользователем одного из: нового голосового действия или второго голосового действия; выбор выбранного голосового действия между новым голосовым действием и вторым голосовым действием содержит: присваивание показателя каждому из: новому

- 5 голосовому действию и второму голосовому действию, и выбор выбранного голосового действия на основе по меньшей мере показателя, присвоенного каждому из: новому голосовому действию и второму голосовому действию; выбор выбранного голосового действия между новым голосовым действием и вторым голосовым действием содержит выбор выбранного голосового действия в ответ на определение, что программное
- 10 приложение, связанное с выбранным голосовым действием, работает на переднем плане.

[0013] В других примерах каждый из этих и других вариантов осуществления могут включать в себя один или более признаков, в том числе: генерация контекстного намерения голосового действия для программного приложения содержит определение,

- 15 что одна или более операций нового голосового действия могут быть выполнены программным приложением; определение, что транскрипция высказывания пользователя, полученная конкретным устройством, имеющим установленное программное приложение, аналогична одному или более триггерным элементам, связанным с контекстным намерением голосового действия, в ответ на это определение,
- 20 предоставление системой голосовых действий конкретному устройству данных, указывающих запрос на пользовательский ввод, который подтверждает, указывало ли высказывание пользователя один или более триггерных элементов или должно было вызвать выполнение программным приложением нового голосового действия, в ответ на этот запрос, прием системой голосовых действий и от конкретного устройства
- 25 данных, указывающих подтверждение, и, в ответ на прием данных, указывающих подтверждение, предоставление системой голосовых действий конкретному устройству контекстного намерения голосового действия, тем самым запроса, чтобы программное приложение, установленное на конкретном устройстве, выполнило одну или более операций нового голосового действия.

- 30 [0014] В других примерах каждый из этих и других вариантов осуществления могут включать в себя один или более признаков, содержащих: прием системой голосовых действий от разработчика, связанного с программным приложением, запроса на развертывание нового голосового действия для программного приложения, и в ответ на этот запрос, развертывание нового голосового действия для программного

- 35 приложения в ответ на этот запрос, при этом развертывание нового голосового действия для программного приложения разрешает новое голосовое действие для программного приложения; прием системой голосовых действий от разработчика, связанного с программным приложением, запроса на отмену развертывания нового голосового действия для программного приложения, и в ответ на этот запрос, отмену развертывания
- 40 нового голосового действия для программного приложения в ответ на этот запрос, при этом отмена развертывания нового голосового действия для программного приложения запрещает новое голосовое действие для программного приложения.

- [0015] В других примерах каждый из этих и других вариантов осуществления могут включать в себя один или более признаков, содержащих: прием системой голосовых действий от разработчика, связанного с программным приложением, запроса на разрешение тестирования нового голосового действия, при этом запрос указывает одно или более устройств, для которых должно быть разрешено новое голосовое действие, и, в ответ на этот запрос, разрешение нового голосового действия для одного или более

указанных устройств, при этом новое голосовое действие запрещено для устройств, которые не включены в указанные устройства; прием системой голосовых действий указания относительно высказывания пользователя, полученного конкретным устройством, имеющим установленное программное приложение, определение системой

5 голосовых действий, что транскрипция высказывания пользователя соответствует одному или более триггерным элементам, связанным с контекстным намерением голосового действия, и в ответ на это определение, предоставление системой голосовых действий конкретному устройству контекстного намерения голосового действия, тем самым запроса, чтобы программное приложение, установленное на конкретном

10 устройстве, выполнило одну или более операций нового голосового действия.

[0016] Подробности одного или более вариантов осуществления предмета изобретения, описанного в этом описании, изложены в прилагаемых чертежах и описании ниже. Другие потенциальные признаки, аспекты и преимущества предмета изобретения станут очевидны из описания, чертежей и формулы изобретения.

15 КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

[0017] Фиг. 1 изображает иллюстративную систему для системы и службы разработки голосовых действий.

[0018] Фиг. 2 изображает иллюстративную систему для генерации новых голосовых действий с использованием системы и службы разработки голосовых действий.

20 [0019] Фиг. 3 изображает иллюстративную систему для использования голосовых действий.

[0020] Фиг. 4 является блок-схемой последовательности операций иллюстративного процесса, связанного с системой и службой разработки голосовых действий.

[0021] Одинаковые ссылочные позиции на различных чертежах обозначают

25 одинаковые элементы.

### ОСУЩЕСТВЛЕНИЕ ИЗОБРЕТЕНИЯ

[0022] Фиг. 1 изображает пример системы 100 голосовых действий. Система 100 голосовых действий обеспечивает платформу и службу, посредством которой разработчик приложений может назначить новые голосовые действия для приложения, 30 которое ранее было установлено на устройствах других пользователей. После этого пользователи приложения, имеющие установленное на их устройствах приложение, могут использовать голосовые команды, чтобы вызвать выполнение приложением конкретной операции, указанной новым голосовым действием.

[0023] Система 100 голосовых действий принимает от терминалов 102a-102n данные, 35 задающие новые голосовые действия, представленные разработчиками 101a-101n приложений, при этом новые голосовые действия предназначены для одного или более различных приложений. Система 100 голосовых действий может определить, является ли каждое новое голосовое действие допустимым голосовым действием. Каждое новое голосовое действие, для которого определено, что оно является допустимым, может 40 быть введено для генерации намерения, соответствующего новому голосовому действию, и намерение, соответствующее новому голосовому действию, может сохраняться в базе 110 данных голосовых действий приложений системы 100 голосовых действий.

[0024] Например, система 100 голосовых действий может иметь механизм (подсистему) проверки, который принимает и проверяет данные, задающие новое голосовое действие.

45 Проверка данных может включать в себя определение, что форматирование данных, задающих голосовое действие, удовлетворяет требованиям, определение, что новое голосовое действие совместимо с указанным приложением, или, в противном случае, определение, что голосовое действие может быть назначено на основе данных, принятых

от разработчика 101a-101n приложений. Механизм указания системы 100 голосовых действий может принять проверенные данные, задающие новое голосовое действие, и может вводить данные для генерации намерения, соответствующего новому голосовому действию. Намерение затем может быть сохранено в базе 110 данных голосовых

5 действий приложений.

[0025] База 110 данных голосовых действий приложений может хранить голосовые действия, которые относятся к активностям или действиям, которые могут быть выполнены в многочисленных различных программных приложениях. Голосовые действия, включенные в базу 110 данных голосовых действий приложений, могут

10 включать в себя встроенные голосовые действия, которые были представлены разработчиками 101a-101n приложений, когда приложение было создано, а также голосовые действия, которые были представлены разработчиками 101a-101n приложений после того, как приложение было создано, для операций, поддерживаемых приложением.

[0026] Голосовые действия также могут включать в себя зависящие от приложения голосовые действия, которые поддерживаются по умолчанию операционной системой. Эти предварительно поддерживаемые голосовые действия могут быть голосовыми действиями, которые могут быть реализованы в приложениях без необходимости представления системе 100 голосовых действий разработчиком 101a-101n приложений. Например, голосовое действие «выйти из приложения» для выхода из задачи или

20 активности, выполняющейся на переднем плане, может быть автоматически доступно для любого приложения и может быть включено в базу 110 данных голосовых действий приложений для каждого приложения, без необходимости разработчику 101a-101n приложений представлять информацию системе 100 голосовых действий для задания голосового действия «выйти из приложения».

25 [0027] В дополнение к базе 110 данных голосовых действий приложений система 100 голосовых действий может включать в себя базу 120 данных голосовых действий операционной системы (OS), которая хранит голосовые действия, которые не связаны ни с каким конкретным приложением или контекстом. Например, голосовое действие «заблокировать телефон», которое вызывает вход устройства в заблокированное

30 состояние, может быть голосовым действием, которое указано в базе 120 данных голосовых действий OS как голосовое действие, которое не связано с конкретным приложением, или которое универсально доступно, то есть не зависит от статуса устройства, когда обнаружено голосовое действие «заблокировать устройство». Как правило, голосовые действия, сохраненные в базе 120 данных голосовых действий OS,

35 являются голосовыми действиями, которые генерируются не на основе задания разработчиками 101a-101n приложений голосовых действий, так как голосовые действия, сохраненные в базе 120 данных голосовых действий OS, являются общими для операционной среды пользовательского устройства, а не любого конкретного приложения, генерируемого сторонними разработчиками 101a-101n приложений.

40 [0028] Система 100 голосовых действий позволяет пользователю 105, имеющему пользовательское устройство 104, обеспечивать произносимый голосовой ввод для их пользовательского устройства 104, чтобы вызвать выполнение действия пользовательским устройством 104 или приложениями, выполняющимися на их пользовательском устройстве 104. Например, пользователь 105, имеющий

45 пользовательское устройство 104, имеет приложение медиапроигрывателя, запущенное на его пользовательском устройстве 104, и он обеспечивает голосовой ввод «play next» во время выполнения приложения медиапроигрывателя на переднем плане пользовательского устройства 104.

[0029] Аудиоданные, соответствующие голосовому вводу, и контекстная информация, указывающая статус пользовательского устройства 104, в том числе статус приложений, запущенных на пользовательском устройстве 104, передаются системе 100 голосовых действий. В частности, для примера, показанного на фиг. 1, аудиоданные,

- 5 соответствующие голосовому вводу «play next», принимаются механизмом 130 распознавания речи системы 100 голосовых действий, и контекстная информация, указывающая статус пользовательского устройства 104, принимается механизмом 140 анализа контекста системы 100 голосовых действий. Контекстная информация может указывать, что приложение медиапроигрывателя выполняется на переднем плане
- 10 пользовательского устройства 104, что приложение медиапроигрывателя в настоящий момент находится в режиме аудиопроигрывателя, и может указывать другую информацию о статусе пользовательского устройства 104 и приложений, установленных на пользовательском устройстве 104.

[0030] Механизм 130 распознавания речи принимает аудиоданные, соответствующие голосовому вводу, генерирует транскрипцию голосового ввода и обеспечивает транскрипцию голосового ввода селектору 150 голосового действия. Механизм 140 анализа контекста принимает контекстную информацию от пользовательского устройства 104 и обрабатывает контекстную информацию для определения релевантной контекстной информации. Например, механизм 140 анализа контекста может 20 анализировать контекстную информацию для идентификации приложений, которые установлены на пользовательском устройстве 104, и метаданные, связанные с каждым из этих приложений, могут указывать доступные голосовые действия для приложения и контекстную информацию, которая релевантна для определения, какое из голосовых действий может быть разрешено для данного статуса оборудования. В некоторых 25 случаях на основе анализа механизм 140 анализа контекста может определить, что необходима дополнительная контекстная информация для идентификации, какие голосовые действия должны быть разрешены для конкретного статуса пользователя устройства 104, и, таким образом, механизм 140 анализа контекста может запросить дополнительную контекстную информацию от пользовательского 30 устройства 104. Механизм 140 анализа контекста отправляет обработанную контекстную информацию селектору 150 голосового действия.

[0031] Например, как показано на фиг. 1, механизм 130 распознавания речи может принять аудиоданные, соответствующие голосовому вводу «play next» и может получить транскрипцию голосового ввода. Механизм 140 анализа контекста принимает 35 контекстную информацию от пользовательского устройства 104, которая указывает, что приложение медиапроигрывателя выполняется на переднем плане, что приложение медиапроигрывателя работает в режиме аудиопроигрывателя, и которая указывает другую информацию о пользовательском устройстве 104. Например, контекстная информация также может указывать, что приложение социальной сети работает в 40 фоновом режиме на пользовательском устройстве 104, и что время работы от батареи пользователя устройства 104 в настоящий момент составляет 50%. Механизм 150 анализа контекста может принять информацию, указывающую, что на пользовательском устройстве 104 установлены как приложение медиапроигрывателя, так и приложение социальной сети, и может определить, что ни приложение медиапроигрывателя, ни 45 приложение социальной сети не разрешают голосовые действия на основе контекстной информации, которая указывает время работы устройства от батареи. Поэтому механизм 140 анализа контекста может отфильтровать контекстную информацию, чтобы указать только то, что приложение медиапроигрывателя работает на переднем плане

пользовательского устройства 104, что приложение медиапроигрывателя находится в режиме аудиопроигрывателя, и что приложение социальной сети работает в фоновом режиме на пользовательском устройстве 104. Отфильтрованная контекстная информация и транскрипция голосового ввода «play next» затем могут быть предоставлены селектору 150 голосового действия.

[0032] Селектор 150 голосового действия принимает транскрипцию голосового ввода от механизма 130 распознавания речи и контекст от механизма 140 анализа контекста, который включает в себя обработанную контекстную информацию. Селектор 150 голосового действия использует транскрипцию и контекст для идентификации

10 конкретного намерения, связанного с конкретным голосовым действием для активации в пользовательском устройстве 104. Например, селектор 150 голосового действия может осуществить доступ к базе 110 данных голосовых действий приложений и базе 120 данных голосовых действий OS для идентификации набора потенциальных голосовых действий, которые разрешены для данного контекста пользовательского устройства 104. Идентификация набора потенциальных голосовых действий может включать в себя идентификацию набора намерений, сохраненных в базе 110 данных голосовых действий приложений или базе 120 данных голосовых действий OS, которые указывают контексты, соответствующие контексту, принятому селектором 150 голосового действия.

[0033] Идентифицировав набор потенциальных голосовых действий, которые 20 разрешены для представленного контекста, селектор 150 голосового действия может сравнить транскрипцию голосового ввода с одной или более триггерными фразами, связанными с каждым из разрешенных голосовых действий. Вообще, триггерная фраза может включать в себя один или более триггерных терминов, она работает как идентификатор для голосового действия, так что обнаружение одного или более 25 терминов конкретной триггерной фразы приводит к идентификации и активации голосового действия, соответствующего конкретной триггерной фразе. Например, селектор 150 голосового действия может сравнить транскрипцию с соответствующими триггерными фразами, связанными с намерениями разрешенных голосовых действий. Селектор 150 голосового действия идентифицирует конкретное намерение для 30 конкретного голосового действия на основе определения, что транскрипция соответствует триггерному термину, указанному намерением, связанным с конкретным голосовым действием.

[0034] Например, селектор 150 голосового действия может принять транскрипцию голосового ввода «play next» и контекст для пользовательского устройства 104 и может 35 использовать принятый контекст для идентификации потенциальных голосовых действий для инициации на пользовательском устройстве 104, то есть голосовых действий, которые могут быть инициированы в пользовательском устройстве 104 на основе текущего статуса пользовательского устройства 104. После определения набора потенциальных голосовых действий для контекста пользовательского устройства 104, селектор 150 40 голосового действия сравнивает транскрипцию голосового ввода «play next» с триггерными фразами, указанными намерениями потенциальных голосовых действий. Селектор 150 голосового действия может определить, что транскрипция «play next» соответствует триггерной фразе, указанной намерением голосового действия «play next» для приложения медиапроигрывателя.

[0035] Транскрипция идентифицируется как соответствующая триггерной фразе на 45 основе определения, что по меньшей мере часть транскрипции соответствует одному или более элементам триггерной фразы. Например, соответствие между транскрипцией и триггерной фразой может быть идентифицировано на основе того, что каждый из

элементов транскрипции соответствует каждому из элементов триггерной фразы. В некоторых реализациях соответствие может быть идентифицировано на основе определения, что менее чем все элементы транскрипции соответствуют элементам триггерной фразы. Например, транскрипция может соответствовать триггерной фразе,

- 5 даже если некоторые из элементов триггерной фразы отсутствуют в транскрипции, если транскрипция включает в себя дополнительные элементы по сравнению с элементами триггерной фразы, или если один или более элементов транскрипции отличаются от элементов триггерной фразы. Таким образом, каждая из транскрипций: «новое электронное письмо», «послать новое электронное письмо» или «открыть новое 10 электронное письмо» может быть идентифицирована как соответствующая триггерной фразе «открыть новое электронное письмо», несмотря на то, что каждая транскрипция не точно соответствует триггерной фразе. В некоторых случаях транскрипция может быть идентифицирована как соответствующая триггерной фразе, если транскрипция включает в себя некоторые или все элементы триггерной фразы в другом порядке,
- 15 отличаясь от порядка, указанного триггерной фразой. Например, транскрипция «заблокировать телефон» может быть идентифицирована как соответствующая триггерной фразе «телефон заблокировать». В некоторых случаях транскрипция может быть идентифицирована как соответствующая триггерной фразе на основе определения, что транскрипция включает в себя один или более синонимов или альтернативных 20 элементов для элементов триггерной фразы. Таким образом, транскрипция «набросать новое электронное письмо» или «открыть новое электронное письмо» может быть идентифицирована как соответствующая триггерной фразе «написать новое электронное письмо» на основе того, что «набросать» является синонимом «написать», и «открыть» является идентифицированным альтернативным элементом для «написать».

- 25 [0036] Идентифицировав конкретное голосовое действие, селектор 150 голосового действия предоставляет данные триггера действия пользовательскому устройству 104, что вызывает инициирование активности или действия, связанного с конкретным голосовым действием. Чтобы сделать это, селектор 150 голосового действия может получить информацию для управления пользовательским устройством 104 так, чтобы 30 выполнить действие или активность, связанную с выбранным голосовым действием. В некоторых случаях управление пользовательским устройством 104 так, чтобы выполнить действие или активность, связанную с выбранным голосовым действием, может включать в себя инициирование намерения выбранного голосового действия. Инициирование намерения выбранного голосового действия может вызвать 35 предоставление пользовательскому устройству 104 информации для инициирования активности или действия, связанного с выбранным голосовым действием.

- [0037] Например, намерение для выбранного голосового действия может включать в себя данные, которые вызывают выполнение приложением, связанным с голосовым действием, действий или активностей, связанных с голосовым действием. Выбранное 40 намерение может быть передано пользовательскому устройству 104, так что получение намерения пользовательским устройством 104 может действовать как запрос или может активировать исполнение действий или активностей, связанных с голосовым действием. Альтернативно, селектор 150 голосового действия может определить другие данные, которые работают как данные триггера действия, которые вызывают выполнение 45 выбранного голосового действия в пользовательском устройстве 104, и может передать эту информацию пользовательскому устройству 104. Например, намерение может только идентифицировать действия или активности, которые должны быть выполнены пользовательским устройством 104, для выполнения голосового действия, и селектор

150 голосового действия может определить данные триггера действия, которые могут управлять приложением на пользовательском устройстве 104 так, чтобы выполнить действия или активности, связанные с голосовым действием. Определенные данные триггера действия могут быть переданы пользовательскому устройству 104, так что

5 данные триггера действия вызывают выполнение действий или активностей, связанных с голосовым действием.

[0038] Например, после выбора голосового действия «play next» из набора потенциальных голосовых действий, селектор 150 голосового действия передает намерение для голосового действия «play next» пользовательскому устройству 104 или

10 получает другую информацию для управления приложением медиапроигрывателя, выполняющимся на пользовательском устройстве 104, чтобы перейти к следующей песне. Селектор 150 голосового действия передает данные для управления приложением медиапроигрывателя так, чтобы перейти к следующей песне, пользовательскому устройству 104, которое в свою очередь инициирует процесс, чтобы приложение

15 медиапроигрывателя, запущенное на пользовательском устройстве 104, перешло к следующей песне, например, перешло к песне «Fame» Дэвида Боуи.

[0039] Говоря кратко, как уже обсуждалось, система фиг. 1 включает в себя один или более терминалов 102a-102n, соответствующих одному или более разработчикам 101a-101n приложений. Каждый из терминалов 102a-102n может осуществлять связь с

20 системой 100 голосовых действий, например, через одну или более проводных или беспроводных сетей. Каждый из терминалов 102a-102n может быть мобильным устройством, таким как сотовый телефон, смартфон, планшетный компьютер, портативный компьютер, персональный цифровой помощник (PDA), нетбук или другое

25 мобильное вычислительное устройство, или может быть любым стационарным вычислительным устройством, таким как настольный компьютер или другой стационарный компьютер. Пользовательское устройство 104, связанное с пользователем 105, также может осуществлять связь с системой 100 голосовых действий через одну или более проводных или беспроводных сетей, при этом пользовательское устройство 104 также может быть мобильным или стационарным вычислительным устройством,

30 таким как сотовый телефон, смартфон, планшетный компьютер, нетбук, персональный цифровой помощник, портативный компьютер, настольный компьютер или другое вычислительное устройство.

[0040] Иллюстративная система 100 голосовых действий, показанная на фиг. 1, включает в себя базу 110 данных голосовых действий приложений, базу 120 данных

35 голосовых действий операционной системы (OS), механизм 130 распознавания речи, механизм 140 анализа контекста и селектор 150 голосового действия. Каждый из компонентов системы 100 голосовых действий, в том числе база 110 данных голосовых действий приложений, база 120 данных голосовых действий OS, механизм 130 распознавания речи, механизм 140 анализа контекста и селектор 150 голосового

40 действия, может осуществлять связь с одним или более другими компонентами системы 100 голосовых действий через один или более проводных или беспроводных путей передачи данных, которые обеспечивают обмен электронными сообщениями. В некоторых реализациях один или более компонентов системы 100 голосовых действий могут быть объединены, так что их функции выполняются одним компонентом, или

45 могут быть представлены двумя или более компонентами, так что их функции распределяются между этими двумя или более компонентами. Компоненты системы 100 голосовых действий могут быть реализованы на одном вычислительном устройстве, таком как одна серверная система, или могут быть реализованы на множестве

вычислительных устройств, которые осуществляют связь через одну или более проводных или беспроводных путей передачи данных, которые обеспечивают обмен электронными сообщениями между компонентами.

[0041] Фиг. 2 изображает иллюстративную систему, которая позволяет разработчику

5 приложений генерировать новые голосовые действия для приложения, которое установлено на одном или более пользовательских устройствах. Например, после того, как приложение было выпущено и загружено на одно или более пользовательских устройств, разработчик приложений может задать новое голосовое действие с использованием службы разработки голосовых действий. Служба разработки голосовых 10 действий может развернуть новое голосовое действие, чтобы обеспечить новое голосовое действие для приложения без необходимости модификаций в самом приложении.

[0042] Говоря кратко, фиг. 2 изображает систему 200 голосовых действий разработчика, которая предоставляет службы разработки голосовых действий разработчику 201 приложений, который связан с терминалом 202. Терминал 202 может 15 осуществлять связь с системой 200 голосовых действий разработчика через одно или более проводных или беспроводных сетевых подключений, таких как проводное или беспроводное подключение к сети Интернет, которое позволяет терминалу 202 обмениваться электронными сообщениями с системой 200 голосовых действий разработчика. Система 200 голосовых действий разработчика включает в себя механизм

20 210 проверки, который осуществляет связь с базой 215 данных критериев проверки. Система 200 голосовых действий разработчика дополнительно включает в себя механизм 220 введения грамматики, который осуществляет связь с серверной частью 225 метаданных приложения и базой 230 данных голосовых действий приложений. База 230 данных голосовых действий приложений может быть аналогична базе 110 данных 25 голосовых действий приложений на фиг. 1 и может хранить намерения, генерируемые из грамматик, которые представляются разработчиками приложений для задания новых голосовых действий для приложений или которые основаны на информации, представленной разработчиками приложений, которая задает новые голосовые действия. В некоторых реализациях система 200 голосовых действий разработчика,

30 предоставляющая службу для разработчиков приложений по назначению новых голосовых действий, может быть частью системы 100 голосовых действий фиг. 1. В некоторых реализациях операции, выполняемые компонентами системы 200 голосовых действий разработчика, могут выполняться другими компонентами системы 200 голосовых действий разработчика. Система 200 голосовых действий разработчика, 35 может содержать больше, меньше или другие компоненты для выполнения операций, описанных как выполняемых системой 200 голосовых действий разработчика.

[0043] Как показано на фиг. 2, разработчик 201 приложений может представить информацию, задающую новое голосовое действие, системе 200 голосовых действий разработчика с использованием терминала 202. Например, терминал 202 может

40 включать в себя приложение для службы голосовых действий разработчика, или терминал 202 может иметь возможность получить доступ к службе голосовых действий разработчика через сетевое соединение, например, путем доступа к службе голосовых действий разработчика на веб-сайте через Интернет. В других реализациях служба голосовых действий разработчика может находиться в системе 200 голосовых действий разработчика, 45 так что терминал 202 может получить доступ к службе голосовых действий разработчика в системе 200 голосовых действий разработчика по сетевому соединению между терминалом 202 и системой 200 голосовых действий разработчика.

[0044] Чтобы реализовать новое голосовое действие для конкретного приложения,

разработчик 201 приложений может представить новую грамматику системе 200 голосовых действий разработчика, которая задает новое голосовое действие, или может представить другую информацию, которая задает новое голосовое действие, системе 200 голосовых действий разработчика. В примере, показанном на фиг. 2, чтобы

- 5 реализовать новое голосовое действие, разработчик 201 приложений сначала задает новое голосовое действие для приложения медиапроигрывателя, которое уже установлено на одном или более пользовательских устройствах. Чтобы сделать это, разработчик 201 приложений представляет грамматику, которая задает новое голосовое действие для приложения медиапроигрывателя. Грамматика, представленная
- 10 разработчиком 201 приложений, указывает приложение, а именно, приложение медиапроигрывателя, триггерную фразу «play next», которая должна быть связана с новым голосовым действием, действие, которое будет выполнено в ответ на обнаружение триггерной фразы «play next», а именно перейти к следующей песне, и контекст для того, когда новое голосовое действие разрешено, а именно, что приложение
- 15 медиапроигрывателя должно работать на переднем плане пользовательского устройства, и что приложение медиапроигрывателя должно быть в режиме аудиопроигрывателя.

[0045] Грамматика, представленная разработчиком 201 приложений, может быть в конкретном формате, который позволяет разработчику 201 приложений легко задать новое голосовое действие, без глубоких знаний о конкретном формате данных,

- 20 используемом системой 200 голосовых действий разработчика для намерений голосового действия. В других реализациях разработчик 201 приложений может представить другую информацию, которая может быть принята механизмом 210 проверки или другим компонентом системы 200 голосовых действий разработчика и преобразована для генерации грамматики, задающей новое голосовое действие. Например, разработчику
- 25 приложений 201 может быть представлена форма, которая включает в себя поля, которые разработчик 201 приложений может заполнить, чтобы задать новое голосовое действие. Форма может позволять разработчику 201 приложений указать информацию о новом голосовом действии, например, приложение, триггерную фразу для нового голосового действия и действие или активность, которая будет выполняться приложение
- 30 или в отношении приложения в ответ на триггерную фразу. Форма может позволять разработчику 201 приложений обеспечивать контекст, который требует конкретного статуса пользовательского устройства или одного или более приложений, работающих на пользовательском устройстве, для того, чтобы новое голосовое действие было разрешено, то есть чтобы голосовое действие было потенциальным голосовым
- 35 действием, которое пользователь может активировать с использованием триггерной фразы. В форму может быть введена другая информация, необходимая или опционально используемая для задания нового голосового действия. Вводимые данные в полях формы, обеспеченные разработчиком 201 приложений, могут быть преобразованы в грамматику для предоставления системе 200 голосовых действий разработчика, или
- 40 вводимые данные могут быть приняты системой 200 голосовых действий разработчика и преобразованы в грамматику в системе 200 голосовых действий разработчика. Альтернативно, разработчик 201 приложений может задать новое голосовое действие путем ввода необходимой информации с использованием формата грамматики, который представляется системе 200 разработчика голосового действия, или с использованием
- 45 другого формата, такого как формат значений, разделенных разделителем (CSV).

[0046] В некоторых реализациях новое голосовое действие может быть разработано для выполнения одной или более операции в программном приложении, которое работает на пользовательском устройстве, для выполнения одного или более операций

в отношении пользовательского устройства вообще (то есть для управления аппаратным обеспечением пользовательского устройства), управления приложением, которое является внешним по отношению к пользовательскому устройству (то есть, приложением, работающим на другом устройстве), или управления другим устройством,

- 5 которое является внешним по отношению к пользовательскому устройству. Например, разработчик 201 приложений может обеспечить грамматику, задающую новое голосовое действие для управления конкретным приложением на пользовательском устройстве (например, приложением социальной сети или приложением медиапроигрывателя), управления пользовательским устройством (например, для изменения ориентации 10 экрана устройства или выключения устройства), для управления приложением, работающим на другом устройстве (например, для управления программным обеспечением кабельного телевидения, работающим на приставке кабельного телевидения, которое является внешним по отношению к пользовательскому устройству), или управления другим устройством, которое является обособленным от 15 пользовательского устройства (например, для управления дверью гаража с использованием голосового ввода, предоставленного пользовательскому устройству). В некоторых реализациях голосовое действие может быть задано так, что активация голосового действия вызывает выполнение двух или более различных действий или активностей в двух или более различных приложениях или устройствах.

- 20 [0047] Контекст, указанный разработчиком 201 приложений, идентифицирует одно или более условий, которые должны быть удовлетворены для того, чтобы голосовое действие было разрешено, то есть такое, что голосовое действие не будет активировано триггерной фразой, связанной с голосовым действием, если также не будет удовлетворен контекст. Разработчик 201 приложений может задать контекст как имеющий одно или 25 более различных требований или условий.

- [0048] В некоторых случаях контекст, представленный разработчиком 201 приложений, может указывать статус или атрибуты пользовательского устройства, такие как включено ли пользовательское устройство или заблокировано, имеет ли пользовательское устройство камеру, гироскоп, барометр или другой компонент или 30 функциональную возможность. Таким образом, например, голосовое действие сделать снимок с использованием приложения, установленного на пользовательском устройстве, не может быть разрешено, если контекстная информация, принятая из пользовательского устройства, указывает, что у пользовательского устройства нет камеры.

- [0049] В других примерах контекст может требовать, чтобы конкретные приложения 35 работали на переднем плане или в фоновом режиме пользовательского устройства. Например, как показано на фиг. 2, голосовое действие «play next» может быть разрешено только тогда, когда приложение медиапроигрывателя работает на переднем плане пользовательского устройства 204, но не когда приложение медиапроигрывателя закрыто или работает в фоновом режиме пользовательского устройства 204. В других 40 примерах голосовое действие может быть задано с контекстом, который разрешает голосовое действие, когда приложение работает в фоновом режиме. Например, приложение социальной сети может иметь связанное голосовое действие на одобрение запроса на соединение с другим пользователем, и пользователь может иметь возможность активировать голосовое действие на одобрение запроса на соединение с 45 другим пользователем, даже если приложение социальной сети работает только в фоновом режиме пользовательского устройства.

[0050] Контекст может дополнительно или альтернативно требовать, чтобы конкретное приложение работало в конкретном режиме, чтобы голосовое действие

было разрешено. Режим может быть конкретной активностью или задачей, которую исполняет приложение. Например, может быть определено, что почтовая программа находится в режиме написания электронного письма, приложение медиапроигрывателя может быть в режиме аудиопроигрывателя, или приложение камеры может быть в

5 режиме камеры или режиме просмотра фотоальбома. Пользовательское устройство может иметь возможность определять, в каком из режимов работает конкретное приложение, и может включать эту информацию в контекстную информацию, которая используется для определения, разрешено ли конкретное голосовое действие.

[0051] Дополнительно, в некоторых случаях контекст может требовать, чтобы режим

10 приложения имел некоторый статус. Например, контекст может указывать, что голосовое действие «пауза» фильма может быть включено только тогда, когда приложение медиапроигрывателя находится в режиме видеопроигрывателя, и когда статус приложения в режиме видеопроигрывателя, то это означает, что приложение в

15 настоящий момент проигрывает фильм. В некоторых случаях пользовательское устройство может не иметь возможности определить статус приложения, работающего в конкретном режиме. В этих случаях может быть необходимо сконфигурировать приложение так, чтобы оно предоставляло информацию, указывающую статус

приложения в конкретном режиме. Таким образом, указание такого условия в контексте для конкретного приложения может требовать модификации самого приложения, чтобы

20 контекстная информация, обеспеченная пользовательским устройством, включала в себя требуемую информацию о статусе.

[0052] В некоторых примерах контекст, указанный для голосового действия, может требовать, чтобы конкретные объекты или типы объектов были отображены на экране пользовательского устройства, чтобы разрешить голосовое действие. Например,

25 контекст может указывать, что голосовое действие «сохранить изображение» разрешается, только если изображение выводится на дисплее пользовательского устройства, или может указывать, что голосовое действие «выделить» доступно, только если на дисплее представлен текст. В другом примере голосовое действие по выбору конкретного элемента в списке, такое как голосовое действие «выбрать первый» может

30 быть доступно, только если контекстная информация указывает, что номер «1» или пункт маркированного списка представлен на дисплее пользовательского устройства, или если есть список элементов, представленный на дисплее. В некоторых случаях пользовательское устройство может не иметь возможности определения, какая

35 информация представлена на его дисплее в данный момент, например, когда дисплеем управляет приложение. Таким образом, если контекст указывает, что определенная информация или типы информации должны выводиться на дисплей для того, чтобы голосовое действие было разрешено, то может быть необходимо модифицировать приложение так, чтобы оно предоставляло эту информацию. Пользовательское устройство затем может включить информацию, указывающую, что выводится на

40 дисплее пользовательского устройства, в контекстную информацию, которая используется для определения, должно ли быть разрешено конкретное голосовое действие.

[0053] В других случаях контекст может требовать, чтобы уведомление или конкретный тип уведомления было выведено на пользовательском устройстве.

45 Например, почтовая программа, которая обеспечивает всплывающие уведомления для пользователей, когда получено новое электронное письмо, может иметь связанное голосовое действие «прочитать электронное письмо», которое разрешено, когда уведомление о новом электронном письме выводится приложением на дисплее

пользовательского устройства. Таким образом, контекстная информация может указывать, выводится ли уведомление на дисплее пользовательского устройства, и эта информация может использоваться в определении, должно ли быть разрешено голосовое действие «прочитать электронное письмо».

- 5 [0054] В некоторых примерах контекст может требовать, чтобы пользовательское устройство было в конкретном географическом местоположении для того, чтобы голосовое действие было разрешено. Пользовательское устройство может иметь возможность определения его географического местоположения с использованием триангуляции по сотовым вышкам, на основе получения доступа к интернет-соединению,
- 10 которое связано с конкретным географическим местоположением, с использованием технологии глобальной навигационной системы (GPS) или используя другое средство. Пользовательское устройство может включать информацию, указывающую его географическое местоположение, в контекстную информацию, и голосовое действие может быть разрешено на основе того, что географическое местоположение
- 15 удовлетворяет требованию контекста к географическому местоположению. В качестве примера, контекст, связанный с голосовым действием для приложения оператора розничной торговли, может указывать, что некоторое голосовое действие должно быть обработано, только если пользовательское устройство, имеющее приложение оператора розничной торговли, находится в пределах одного из магазинов оператора розничной
- 20 торговли. Голосовое действие для приложения оператора розничной торговли может быть разрешено на основе контекстной информации от пользовательского устройства, указывающей, что географическое местоположение пользовательского устройства соответствует географическому местоположению, которое находится в пределах одного из магазинов оператора розничной торговли.
- 25 [0055] Контекст, заданный разработчиком приложений, также может указывать, что голосовое действие, связанное с контекстом, разрешено только в течение некоторого времени, дат, месяцев, сезонов, или когда удовлетворены другие условия по времени. Например, контекст может указывать, что некоторое голосовое действие разрешено, только если контекстная информация, принятая от пользовательского устройства или
- 30 из другого источника информации, удовлетворяет условию по времени. Например, голосовое действие для банковского приложения может быть связано с контекстом, который включает в себя условие по времени, так что голосовое действие разрешено, только если контекстная информация указывает, что время, когда голосовое действие представлено, находится между 6:00 и 22:00.
- 35 [0056] Контекст может указывать другие требования для голосового действия, которое должно быть разрешено. Например, система 200 голосовых действий разработчика может иметь доступ к информации из источников помимо пользовательского устройства или может быть иметь возможность принимать информацию от пользовательского устройства, которая получена из одного или более
- 40 источников информации. Информация, принятая из других источников, может быть контекстной информацией, требуемой для разрешения голосового действия. Такая информация может включать в себя, например, информацию о погоде, эмоциональную информацию для пользователя, информацию о новостях, информацию о фондовом рынке или другую информацию. Например, система 200 голосовых действий
- 45 разработчика может иметь возможность осуществлять доступ к другим источникам информации через одно или более проводных или беспроводных сетевых соединений, например, интернет-соединение или другое сетевое соединение к серверной системе. Контекст для голосового действия может указывать, что голосовое действие разрешено,

только если информация о погоде для географической области, соответствующей местоположению соответствующего пользовательского устройства, указывает, что в этом местоположении идет дождь. Чтобы проверить, выполняется ли условие контекста, система 200 голосовых действий разработчика может иметь возможность получать 5 доступ к информации о погоде для известного местоположения пользовательского устройства через одно или более сетевых соединений.

[0057] Грамматика или другая информация, задающая новое голосовое действие, представленная разработчиком 201 приложением, может быть представлена системе 200 голосовых действий разработчика из терминала 202 по сети. Представленная 10 грамматика принимается в системе 200 голосовых действий разработчика механизмом 210 проверки. Если разработчик 201 приложений представляет информацию системе 200 голосовых действий разработчика, которая задает голосовое действие, но которая находится не в формате грамматики, система 200 голосовых действий разработчика может преобразовать информацию, задающую новое голосовое действие, в грамматику.

15 Механизм 210 проверки анализирует грамматику для определения, может ли быть введена представленная грамматика в намерение для нового голосового действия. Например, механизм 210 проверки может иметь возможность определять, является ли триггерная фраза, указанная грамматикой, допустимой триггерной фразой, может определять, может ли активность, указанная грамматикой, быть выполнена указанным 20 приложением, может определять, является ли синтаксически допустимым формат представленной грамматики, может определять, является ли допустимым контекст, указанный грамматикой, или может в противном случае определять, является ли представленная грамматика допустимой грамматикой, которая может быть введена для генерации намерения для нового голосового действия. Чтобы выполнить проверку, 25 механизм 210 проверки может осуществлять доступ к базе 215 данных критериев проверки, которая включает в себя правила, используемые для проверки грамматики.

[0058] Например, механизм 210 проверки принимает грамматику, которая указывает новое голосовое действие для приложения медиапроигрывателя. Грамматика указывает 30 триггерную фразу «play next», что триггерная фраза «play next» должна вызывать переход приложения медиапроигрывателя к следующей песне, и контекст, указывающий, что новое голосовое действие должно быть разрешено, когда приложение медиапроигрывателя работает на переднем плане и находится в режиме аудиопроигрывателя. Механизм 210 проверки может осуществить доступ к правилам проверки в базе 215 данных критериев проверки и проверить принятую грамматику 35 на основе правил. Например, механизм 210 проверки может определить, является ли приложение медиапроигрывателя существующим приложением, так что новые голосовые действия могут быть созданы для указанного приложения. Механизм 210 проверки может осуществить доступ к правилу, которое указывает, что триггерная фраза должна состоять из более чем одного произносимого слога по длине, и может 40 признать правильной триггерную фразу «play next» на основе установления, что триггерная фраза «play next» длиннее чем один произносимый слог. Правила проверки могут указывать список возможных действий, которые приложение медиапроигрывателя может выполнять, и могут признавать правильным указанное действие перейти к следующей песне на основе установления, что переход к следующей песне является 45 операцией, которую механизм 210 проверки может выполнить. Механизм 210 проверки может проверять контекст, указанный грамматикой, чтобы убедиться, что контекст не содержит противоречий. Например, механизм проверки может убедиться, что указанный контекст не требует, чтобы приложение работало на переднем плане и также

работало в фоновом режиме для того, чтобы новое голосовое действие было разрешено, так как только одно из этих двух условий может быть удовлетворено. Другие критерии проверки могут быть применены к принятой грамматике для определения, допустима ли грамматика, представленная разработчиком 201 приложений.

- 5 [0059] Другие формы проверки могут быть выполнены механизмом 210 проверки. Например, механизм 210 проверки может определять, разрешено ли разработчику 201 приложений генерировать новые голосовые действия для указанного приложения, может определять, была ли уплачена плата, требуемая для генерации нового голосового действия, или может в противном случае определять, может ли грамматика, 10 представленная разработчиком 201 приложений, привести к созданию намерения для нового голосового действия.

[0060] На основе результата проверки, выполненной механизмом 210 проверки, может быть обеспечен ответ разработчику 201 приложений, который указывает, допустима ли грамматика. Если механизм 210 проверки определяет, что грамматика 15 недопустима, механизм 210 проверки может вызвать вывод терминалом 202 информации, указывающей, что грамматика недопустима, и запрашивающей, чтобы разработчик 201 приложений скорректировал грамматику, или может предложить способы, как разработчик 201 приложений может скорректировать грамматику. Например, если определено, что грамматика недопустима, потому что она указывает действие, которое 20 не может быть выполнено указанным приложением (например, если грамматика указывала, что новое голосовое действие должно вызвать открытие приложением медиапроигрывателя нового электронного письма), то механизм 210 проверки может вызвать вывод в терминале 202 информации, которая указывает, что грамматика указывает действие, которое несовместимо с указанным приложением, и может указать, 25 что разработчик 201 приложений может скорректировать грамматику путем изменения либо указанного действия, либо указанного приложения. Если представленная грамматика допустима, механизм 210 проверки может обеспечить указание, что представленная грамматика допустима.

[0061] Дополнительно, механизм 210 проверки может запросить дополнительный 30 ввод от разработчика 201 приложений. Например, если определено, что грамматика допустима, механизм 210 проверки может представить разработчику 201 приложений возможность продолжить генерацию новой голосовой команды, так что намерение для новой голосовой команды не будет генерироваться из грамматики, если разработчик 201 приложений не подтвердит, что он хотел бы продолжить генерацию новой голосовой 35 команды. Это может позволить разработчику 201 приложений определить, представляет ли он должным образом грамматику системе 200 голосовых действий разработчика, без необходимости продолжать грамматическое введение или развертывание нового голосового действия.

[0062] На основе грамматики, представленной проверенным разработчиком 201 40 приложений и, дополнительно, предоставления разработчиком 201 приложений ввода, указывающего, что процесс генерации голосового действия должен быть продолжен, грамматика может быть представлена механизмом 210 проверки механизму 220 введения грамматики. Механизм 220 введения грамматики может вводить принятую грамматику для генерации намерения для нового голосового действия. Сгенерированное 45 намерение может указывать информацию, включенную в грамматику, задающую новое голосовое действие. Например, механизм 220 введения грамматики может вводить принятую грамматику для генерации экземпляра структуры данных намерения, которая указывает одно или более приложений, триггерную фразу, действия или активности

для выполнения в ответ на обнаружение триггерной фразы и контекст, указанный для нового голосового действия. Механизм 220 введения грамматики может сохранить намерение для нового голосового действия в базе 230 данных голосовых действий приложений.

- 5 [0063] Например, как показано на фиг. 2, механизм 220 введения грамматики может принять грамматику, представленную разработчиком 201 приложений, которая была проверена механизмом 210 проверки. Механизм 220 введения грамматики может сгенерировать новое намерение, «Намерение Y», для нового голосового действия на основе принятой грамматики. Для генерации намерения механизм 220 введения
- 10 грамматики может сгенерировать новый экземпляр структуры данных намерения и может заполнить структуру данных намерения с помощью информации, содержащейся в принятой грамматике, включающей в себя информацию, указывающую приложение медиапроигрывателя, информацию, указывающую триггерную фразу «play next», информацию, указывающую, что голосовое действие «play next» должно быть связано
- 15 с переходом к следующей песне в приложении медиапроигрывателя, и информацию, указывающую контекст, указывающий что голосовое действие «play next» должно быть разрешено, когда приложение медиапроигрывателя работает на переднем плане и работает в режиме аудиопроигрывателя.

- 16 [0064] Новое «Намерение Y» может быть сохранено механизмом 220 введения грамматики в базе 230 данных голосовых действий приложений, которая включает в себя другие намерения для других голосовых действий, такие как «Намерение X» и «Намерение Z». Каждое из «Намерения X» и «Намерения Z» может быть связано с другими голосовыми действиями для других приложений или с другими голосовыми действиями для приложения медиапроигрывателя. Как только новое «Намерение Y»
- 25 было сохранено в базе 230 данных голосовых действий приложений, новое голосовое действие может быть развернуто, так что новое голосовое действие может использоваться пользователями, имеющими сохраненное на их устройствах приложение медиапроигрывателя. То есть после развертывания пользователи приложения медиапроигрывателя могут использовать новое голосовое действие «play next» без необходимости обновлять их приложения.

- 26 [0065] В некоторых реализациях введение грамматики для генерации намерения также может включать в себя генерацию одного или более альтернативных триггерных элементов, которые основаны по меньшей мере на триггерном элементе, представленном разработчиком 201 приложений, и включение альтернативных триггерных элементов
- 35 в намерение, генерируемое для нового голосового действия. Например, одно или более правил расширения, правил синонимов, правил опционализации, правил подстановки, правил замены или других правил могут быть применены к триггерному элементу, представленному разработчиком 201 приложений, для определения альтернативных триггерных элементов. Механизм 220 введения грамматики может включать
- 40 альтернативные триггерные элементы в намерение для нового голосового действия, так что обнаружение триггерного элемента, представленного разработчиком 201 приложений, или альтернативного триггерного элемента может приводить к активации голосового действия.

- 41 [0066] Дополнительно, в некоторых реализациях механизм 220 введения грамматики
- 45 может сохранять информацию, относящуюся к новому намерению, отдельно от структуры данных намерения. Например, информация, указывающая контекст, требуемый для того, чтобы новое голосовое действие было разрешено, или триггерный элемент, связанный с новым голосовым действием, может храниться отдельно от

намерения для нового голосового действия. В таких случаях контекстная информация, принятая от пользовательского устройства, или триггерный элемент, принятый от пользовательского устройства, может сравниваться с контекстом и триггерным элементом, связанными с намерением, без необходимости доступа к данным в намерении.

- 5 Намерение может активироваться на основе установления, что принятая контекстная информация и/или триггерный элемент соответствуют соответствующему контексту и/или триггерному элементу, связанным с намерением. Таким образом, только информация, необходимая для активации активности или действия, связанного с новым голосовым действием, должна быть включена в намерение, а другая информация, 10 используемая для выбора намерения для активации, для идентификации намерения, или другая информация может храниться вне намерения. Например, система 200 голосовых действий разработчика может хранить таблицу поиска, связанный список или другие данные, которые связывают информацию о контексте или триггерном элементе с конкретными намерениями, так что идентификация конкретной связки 15 контекста и триггерного элемента может быть причиной активации намерения, связанного с этой связкой.

[0067] В некоторых реализациях система 200 голосовых действий разработчика также может обеспечивать разработчику 201 приложений возможности тестирования нового голосового действия, прежде чем оно будет развернуто для общего пользования.

- 20 Например, система 200 голосовых действий разработчика может обеспечивать разработчику 201 приложений возможности тестирования нового голосового действия с использованием его терминала 202 при отсутствии доступа к голосовому действию для других пользователей. Система 200 голосовых действий разработчика также может обеспечивать разработчику 201 приложений возможности указания одного или более 25 пользовательских устройств, для которых новое голосовое действие должно быть разрешено для тестирования. Например, разработчик 201 приложений может указать, что новое голосовое действие должно быть развернуто для бета-тестировочной группы известных пользовательских устройств, чтобы обеспечить надлежащую работу нового голосового действия до развертывания нового голосового действия для широкой 30 публики.

[0068] В некоторых реализациях система 200 голосовых действий разработчика может позволять тестирование нового голосового действия путем изменения настроек, связанных с намерением, так что намерение может быть идентифицировано и активировано на основе приема голосовой команды от одного из пользовательских

- 35 устройств, указанных для тестирования. Например, система 200 голосовых действий разработчика может хранить параметр для каждого из намерений, сохраненных в базе 230 данных голосовых действий приложений, где параметр указывает, было ли намерение для конкретного голосового действия развернуто для общего пользования, или может идентифицировать конкретные пользовательские устройства, на которых 40 намерение было развернуто для тестирования. После сохранения намерения в базе 230 данных голосовых действий приложений, система 200 голосовых действий разработчика может принять информацию от терминала 202, указывающую, что намерение должно быть разрешено для конкретного набора из одного или более пользовательских устройств для тестирования. В ответ может быть обновлен параметр для намерения, 45 чтобы развернуть намерение на указанном наборе пользовательских устройств, так что голосовое действие, связанное с намерением, может быть активировано. После разрешения для тестирования голосовой ввод, принятый от пользовательского устройства в наборе, может вызвать активацию намерения, тем самым позволяя

тестировать голосовое действие.

[0069] В других вариантах осуществления системы 200 голосовых действий разработчика может позволять тестирование нового голосового действия до сохранения в базе 230 данных голосовых действий приложений намерения для нового голосового

5 действия. Например, разработчик 201 приложений может указать, что он хотел бы протестировать новое голосовое действие, прежде чем его развернуть, и может указать ряд пользовательских устройств или учетных записей пользователей, для которых новое голосовое действие должно быть временно разрешено для тестирования. После генерации намерения для нового голосового действия система 200 голосовых действий

10 разработчика может разрешить голосовое действие для указанных пользовательских устройств или учетных записей пользователей, так что указанное пользовательское устройство или пользовательское устройство, связанное с указанной учетной записью пользователя, может активировать новое голосовое действие, если пользователь обеспечит триггерную фразу, связанную с новым голосовым действием, в качестве

15 ввода одному из пользовательских устройств.

[0070] В других реализациях система 200 голосовых действий разработчика может обеспечивать возможности тестирования новых голосовых действий, прежде чем грамматика для нового голосового действия будет введена механизмом 220 введения грамматики. Например, после проверки принятой грамматики механизм 210 проверки

20 или другой компонент системы 200 голосовых действий разработчика может иметь возможность моделирования работы нового голосового действия, заданного грамматикой, чтобы позволить разработчику 201 приложений или пользователям других устройств протестировать новое голосовое действие. В некоторых случаях моделирование нового голосового действия может требовать использования терминалом

25 202 или другими пользовательскими устройствами, используемыми для тестирования, средства моделирования или режима тестирования, обеспеченного системой 200 голосовых действий разработчика, для моделирования работы нового голосового действия. Обеспечение возможностей тестирования до введения новой грамматики, может обеспечить более быстрое итеративное тестирование новых голосовых действий

30 разработчиком 201 приложений, тем самым ускоряя общий процесс разработки голосового действия.

[0071] Также системой 200 голосовых действий разработчика могут быть обеспечены возможности развертывания голосовых действий и отмены развертывания голосовых действий. Например, после того, как грамматика, задающая новое голосовое действие,

35 была представлена системе 200 голосовых действий разработчика разработчиком 201 приложений, проверена, введена для генерации нового намерения, и новое намерение было сохранено в базе 230 данных голосовых действий приложений, система 200 голосовых действий разработчика может запросить, хочет ли разработчик 201 приложений развернуть новое голосовое действие. Разработчик 201 приложений может

40 выбрать развертывание нового голосового действия, при этом развертывание нового голосового действия позволяет обеспечить активацию нового голосового действия в пользовательских устройствах широкой публики, на которых установлено приложение, связанное с новым голосовым действием. Разработчик 201 приложений может обеспечить ввод в терминале 202, чтобы развернуть новое голосовое действие. Новое

45 голосовое действие затем может быть развернуто, так что новое голосовое действие может быть активировано на других пользовательских устройствах, на которых установлено приложение, связанное с новым голосовым действием. То есть после того, как новое голосовое действие было развернуто, ввод триггерной фразы для нового

голосового действия в пользовательском устройстве с установленным приложением вызовет активацию намерения, связанного с новым голосовым действием. Активация намерения, связанного с новым голосовым действием, вызывает передачу данных намерения или других данных триггера действия пользовательскому устройству, так

- 5 что данные в намерении или данные активации действия вызывают выполнение пользовательским устройством действия или активности голосового действия. Как обсуждалось выше, развертывание намерения может требовать лишь обновления системой 200 голосовых действий разработчика параметра, связанного с намерением нового голосового действия, чтобы сделать намерение потенциальным намерением,
- 10 которое должно активироваться в ответ на голосовой ввод.

[0072] После того, как голосовое действие было развернуто, разработчик 201 приложений может отменить развертывание нового голосового действия, так что новое голосовое действие больше не может быть активировано в пользовательских устройствах, на которых установлено приложение, связанное с новым голосовым действием. Например, разработчик 201 приложений может обеспечить ввод для системы 200 голосовых действий разработчика, идентифицирующий голосовое действие и указывающий, что его намерение состоит в том, чтобы отменить развертывание голосового действия. В ответе системы 200 голосовых действий разработчика может отменить развертывание голосового действия, например, путем обновления параметра, связанного с намерением для голосового действия, так что намерение для голосового действия не может быть активировано. Так как развертывание или отмена развертывания голосового действия требует лишь обновления системой 200 голосовых действий разработчика информации, хранящейся в системе 200 голосовых действий разработчика, без изменения приложения, установленного на устройствах пользователей,

25 разработчик 201 приложений может с легкостью развернуть или отменить развертывание голосовых действий.

[0073] В некоторых реализациях система 200 голосовых действий разработчика также может генерировать ознакомительные примеры для нового голосового действия. Ознакомительный пример может включать в себя информацию, которая может быть представлена пользователю пользовательского устройства, которая сообщает пользователю, что доступны конкретные голосовые действия, возможно в том числе новые голосовые действия, которые были представлены разработчиком после того, как приложение было установлено. В некоторых случаях, ознакомительный пример может быть выведен в ответ на статус пользовательского устройства, который соответствует контексту, связанному с голосовым действием, или в ответ на вопрос пользователя, какие голосовые действия могут быть доступны.

[0074] Например, на основе приема грамматики, задающей новое голосовое действие, механизм 220 введения грамматики или другой компонент системы 200 голосовых действий разработчика могут генерировать ознакомительные примеры для голосового действия, заданного грамматикой, и может сохранять ознакомительные примеры в серверной части 225 метаданных приложения. Ознакомительные примеры могут быть обеспечены для пользовательских устройств, на которых установлено приложение, связанное с новым голосовым действием. Например, метаданные приложения могут периодически обновляться для включения текущего списка ознакомительных примеров для приложения, которые хранятся в серверной части 225 метаданных приложения. Альтернативно, на основе представления пользовательским устройством контекстной информации, которая указывает его текущее статус, пользовательскому устройству может быть передана информация, которое вызывает вывод в пользовательском

устройстве ознакомительных примеров для голосовых действий, разрешенных для этого статуса пользовательского устройства.

[0075] Например, в дополнение к генерации нового «Намерения Y» для голосового действия «play next» система 200 голосовых действий разработчика может сгенерировать

5 один или более ознакомительных примеров для голосового действия «play next».

Ознакомительные примеры могут включать в себя информацию, которая выводится на пользовательских устройствах, которые имеют приложение медиапроигрывателя. Ознакомительный пример может выводиться в пользовательском устройстве на основе удовлетворения контексту, связанному с голосовым действием «play next», или на основе

10 других условий. Например, ознакомительным примером для голосового действия «play next» может быть сообщение «Попробуйте сказать 'play next'», которое выводится всякий раз, когда приложение медиапроигрывателя работает на переднем плане

пользовательского устройства и находится в режиме аудиопроигрывателя. В некоторых реализациях ознакомительный пример может выводиться в ответ на другие события

15 или условия. Например, сообщение ознакомительного примера «Попробуйте сказать 'Play next'», может выводиться в ответ предоставление пользователем пользовательского устройства ввода для пользовательского устройства путем нажатия кнопки или элемента управления или произнесения «Какие голосовые действия я могу произнести?»

[0076] Вообще, ознакомительные примеры для голосовых действий приложения

20 могут быть обеспечены для пользователей, не требуя действия со стороны самих приложений. Например, система 200 голосовых действий разработчика может предоставить данные, включающие в себя ознакомительные примеры,

пользовательскому устройству, и пользовательское устройство может вызвать вывод ознакомительных примеров без необходимости выполнения операций приложениями,

25 относящимся к ознакомительным примерам. В некоторых реализациях OS, работающая на пользовательском устройстве, может иметь возможность приема ознакомительных примеров или доступа к метаданным приложений, которые указывают ознакомительные примеры, и может вызывать вывод в пользовательском устройстве ознакомительных примеров. Таким образом, ознакомительные примеры для голосовых действий,

30 относящиеся к конкретному приложению, могут быть обеспечены для вывода без модификаций конкретного приложения.

[0077] В некоторых реализациях ознакомительные примеры для голосовых действий приложения могут выводиться таким образом, что пользователь понимает, что ознакомительные примеры обеспечиваются не самим приложением. Например,

35 пользовательское устройство может обеспечивать ознакомительные примеры таким образом, что является очевидным, что ознакомительные примеры обеспечиваются системой 200 голосовых действий разработчика или OS, а не конкретными приложениями, относящимися к голосовым действиям. Например, пользовательское устройство может использовать конкретный компьютеризированный тон голоса,

40 распознаваемый как связанный с OS пользовательского устройства, при обеспечении ознакомительных примеров. Альтернативно, ознакомительные примеры могут выводиться таким образом, что понятно, что приложение не обеспечивает ознакомительные примеры. Например, может выводиться ознакомительный пример, который говорит «Попробуйте сказать 'play next' или 'pause', и я сделаю так, что

45 медиапроигрыватель обработает ваш запрос», чтобы отличить роль OS от действий, выполняемых самим приложением медиапроигрывателя. Также могут быть реализованы другие способы указания, что ознакомительные примеры обеспечиваются отдельной системой или приложением.

[0078] Фиг. 3 изображает иллюстративную систему 300 голосовых действий для обработки голосовых действий, принятых от пользователя пользовательского устройства. В некоторых реализациях голосовые действия, обрабатываемые системой 300 голосовых действий, могут быть голосовыми действиями, которые разработчик 5 приложений создал для приложения после того, как приложение было загружено на пользовательское устройство. Голосовое действие, созданное разработчиком приложений, могло быть создано для приложения с использованием системы 200 голосовых действий разработчика на фиг. 2.

[0079] Говоря кратко, система 300 голосовых действий осуществляет связь с

10 пользовательским устройством 304, принадлежащим пользователю 305, через одну или более проводных или беспроводных сетей, таких как одно или более проводных или беспроводных Интернет-соединений или сотовых соединений для передачи данных, которые позволяют пользовательскому устройству 304 обмениваться электронными сообщениями с системой 300 голосовых действий. Система 300 голосовых действий 15 включает в себя механизм 330 распознавания речи и механизм 340 анализа контекста, который обрабатывает информацию, принятую от пользовательского устройства 304 через одно или более сетевых соединений. Система голосовых действий включает в себя средство сопоставления (сопоставитель) 350, которое осуществляет связь с механизмом 330 распознавания речи и механизмом 340 анализа контекста и которое 20 также имеет доступ к базе 310 данных голосовых действий приложений и базе 320 данных голосовых действий OS. Сопоставитель 350 осуществляет связь с блоком 360 создания триггера действия, который имеет доступ к базе 370 данных для хранения данных приложения. Механизм 380 устранения неоднозначности, который имеет доступ 25 к истории 390 активности пользователя, также может посыпать или принимать электронные сообщения от пользовательского устройства 304 через одно или более сетевых соединений.

[0080] Для обработки голосового ввода, представленного пользователем 305, аудиоданные, соответствующие голосовому вводу, и контекстная информация, 30 указывающая статус пользовательского устройства 304 и/или приложений, работающих на пользовательском устройстве 304, предоставляются системе 300 голосовых действий. В некоторых случаях пользовательское устройство 304 может получить голосовой ввод в ответ на выбор пользователем 305 элемента управления, что вызывает старт записи пользовательским устройством 304 речи пользователя, в ответ произнесение пользователем 305 фразы, обнаружение которой вызывает получение пользовательским 35 устройством 304 последующего голосового ввода от пользователя 305, или может по иным причинам решить записать голосовой ввод пользователя 305. Требование обеспечить конкретный ввод для активации записи речи пользователя может предотвратить ненужную обработку пользовательским устройством 304 речи 40 пользователя, то есть сократить количество аудиоданных, которые должны быть обработаны, или предотвратить запись такого большого количества речи пользователя, которое посягает на конфиденциальность пользователя 305, то есть записывать только конкретную речь, которую пользователь 305 обращает к пользовательскому устройству 304.

[0081] В примере, показанном на фиг. 3, например, пользователь 305 может выбрать 45 элемент управления на пользовательском устройстве 304, чтобы вызвать начало записи пользовательским устройством 304 голосового ввода пользователя 305, и после того, как пользователь 305 выбрал элемент управления, пользователь 305 может обеспечить голосовой ввод путем произнесения «play next». Пользовательское устройство 304

может передать аудиоданные, соответствующие голосовому вводу, системе 300 голосовых действий через одну или более сетей. Дополнительно, пользовательское устройство 304 может определить контекстную информацию для пользовательского устройства 304 или приложений, работающих на пользовательском устройстве 304, и контекстная информация для пользовательского устройства 304 может быть передана системе 300 голосовых действий через одну или более сетей.

[0082] В некоторых случаях контекстная информация, переданная от пользовательского устройства 304 системе 300 голосовых действий, включает в себя конкретную информацию, которая идентифицируется информацией, сохраненной на пользовательском устройстве 304. Например, пользовательское устройство 304 может хранить информацию, которая указывает конкретную информацию о статусе, для передачи системе 300 голосовых действий как контекстной информации. Контекстная информация, обеспеченная системе 300 голосовых действий пользовательским устройством 304, может включать в себя стандартизированное тело контекстной информации, так что все пользовательские устройства предоставляют одинаковую контекстную информацию системе 300 голосовых действий, или контекстная информация, предлагаемая пользовательским устройством 304, может быть конкретным набором контекстной информации, назначенной пользовательскому устройству 304. В некоторых реализациях система 300 голосовых действий может запросить некоторую контекстную информацию от пользовательского устройства 304, и контекстная информация, переданная пользовательским устройством, может быть информацией, предоставленной системе 300 голосовых действий в ответ на запрос. В других случаях информация, связанная с приложениями, установленными на пользовательском устройстве 304, такая как метаданные приложений, может указывать контекстную информацию для предоставления системе 300 голосовых действий, и пользовательское устройство 304 может передавать контекстную информацию системе 300 голосовых действий, которая включает в себя информацию, указанную информацией приложения.

[0083] Например, в примере, показанном на фиг. 3, контекстная информация, передаваемая от пользовательского устройства 304 системе 300 голосовых действий, может включить в себя информацию, указывающую, что приложение медиапроигрывателя в настоящий момент работает на переднем плане пользовательского устройства 304, что приложение медиапроигрывателя работает в режиме аудиопроигрывателя, и что текущее время работы от батареи пользовательского устройства 304 составляет 50%.

[0084] Аудиоданные, переданные пользовательским устройством 304, могут быть приняты механизмом 330 распознавания речи системы 300 голосовых действий, и механизм 330 распознавания речи может сгенерировать транскрипцию голосовых аудиоданных. Например, как обсуждалось в отношении механизма 130 распознавания речи на фиг. 1, механизм 330 распознавания речи может иметь возможность приема аудиоданных и генерации транскрипции аудиоданных, соответствующих голосовому вводу пользователя 305. После генерации транскрипции аудиоданных механизм 330 распознавания речи может обеспечить транскрипцию устройству 350 сопоставления.

[0085] В некоторых реализациях в дополнение к приему аудиоданных механизм 330 распознавания речи также может принимать контекстную информацию, предоставленную системе 300 голосовых действий пользовательским устройством 304. Система 330 распознавания речи может настраивать распознавание речи на основе принятой контекстной информации. В некоторых реализациях настройка распознавания

речи на основе контекстной информации может включать в себя коррекцию вероятности, что обнаружены некоторые слова в аудиоданных. Например, система 330 распознавания речи может корректировать вероятности, что обнаружены некоторые слова, на основе их релевантности контексту, указанному контекстной информацией.

- 5 [0086] В качестве примера, механизм 330 распознавания речи может принять аудиоданные, соответствующие произнесению «play next», а также контекстную информацию, указывающую, что приложение медиапроигрывателя в настоящий момент выполняется на переднем плане пользовательского устройства 304. Механизм 330 распознавания речи может настроить распознавание речи на основе контекстной 10 информации так, что механизм 330 распознавания речи генерирует транскрипцию «play next» вместо транскрипции «planets» на основе увеличения механизмом 330 распознавания речи вероятности, что слово «play» или фраза «play next» будут обнаружены в аудиоданных, или уменьшения вероятности обнаружения слова «planets» на основе установления, что эти слова имеют отношение или не имеют отношения к 15 приложению медиапроигрывателя, которое выполняется на переднем плане пользовательского устройства 304.

[0087] Механизм 340 анализа контекста системы голосовых действий принимает контекстную информацию, переданную пользовательским устройством 304, и обрабатывает контекстную информацию для определения контекста пользовательского 20 устройства 304 или приложений, работающих на пользовательском устройстве 304. Аналогично механизму 140 анализа контекста на фиг. 1, механизм 340 анализа контекста системы 300 голосовых действий может обрабатывать принятую контекстную информацию для определения частей принятой контекстной информации, которые относятся к идентификации голосового действия, которое должно быть активировано, 25 или может обрабатывать принятую контекстную информацию для определения, есть ли другая контекстная информация, которая необходима от пользовательского устройства 304 для идентификации голосового действия, которое должно быть активировано. Если необходима дополнительная контекстная информация для определения голосового действия, которое должно быть выполнено, механизм 340 30 анализа контекста может запросить дополнительную контекстную информацию от пользовательского устройства 304, так что пользовательское устройство 304 может ответить на запрос дополнительной контекстной информацией. В некоторых реализациях механизм 340 анализа контекста может модифицировать формат принятой контекстной информации для генерации контекста, который может использоваться системой 300 35 голосовых действий при идентификации голосового действия, которое должно быть активировано. Механизм 340 анализа контекста предоставляет обработанную контекстную информацию устройству 350 сопоставления.

[0088] В примере, показанном на фиг. 3, например, механизм 340 анализа контекста принимает контекстную информацию от пользовательского устройства 304. Принятая 40 контекстная информация может указывать, что приложение медиапроигрывателя выполняется на переднем плане пользовательского устройства 304, что приложение медиапроигрывателя работает в режиме аудиопроигрывателя, и что время работы от батареи пользовательского устройства 304 составляет 50%. Механизм 340 анализа контекста может определить, что информация о времени работы от батареи не важна 45 для определения, какое голосовое действие активировать в ответ на голосовой ввод пользователя 305, и, таким образом, механизм 340 анализа контекста может подать контекст в сопоставитель 350, который указывает только, что приложение медиапроигрывателя работает на переднем плане пользовательского устройства 304

и находится в режиме аудиопроигрывателя.

[0089] Сопоставитель 350 принимает контекст от механизма 340 анализа контекста и транскрипцию от механизма 330 распознавания речи и использует принятый контекст и транскрипцию для идентификации одного или более потенциальных голосовых

5 действий, которые должны быть активированы. Например, сопоставитель 350 может сопоставить контекст, принятый от механизма 340 анализа контекста, с контекстом, связанным с одним или более намерениями, которые включены в базу 310 данных голосовых действий приложений и/или базу 320 данных голосовых действий OS. На основе установления, что принятый контекст соответствует контекстам, связанным с  
10 одним или более намерениями, сопоставитель 350 может идентифицировать эти намерения как потенциальные намерения для активации. После идентификации потенциальных намерений на основе соответствия принятого контекста контекстам, связанным с потенциальными намерениями, сопоставитель 350 может сузить набор потенциальных намерений путем сравнения принятой транскрипции с триггерными фразами, связанными  
15 с потенциальными намерениями. Устройство 350 сопоставления может предоставить информацию, идентифицирующую суженный набор потенциальных намерений, блоку 360 создания триггера действия.

[0090] В некоторых случаях сопоставитель 350 может принять множество

20 потенциальных транскрипций голосового ввода и может идентифицировать суженный набор потенциальных намерений на основе установления, что каждое из потенциальных намерений связано с контекстом, который соответствует контексту, принятому от механизма 340 анализа контекста и одной или более потенциальными транскрипциями голосового ввода, принятым от механизма 330 распознавания речи. Например, механизм 330 распознавания речи может определить множество потенциальных транскрипций  
25 для голосового ввода и может предоставить информацию, указывающую каждую потенциальную транскрипцию, сопоставителю 350. Сопоставитель 350 затем может идентифицировать суженный набор потенциальных намерений, каждый из которых указывает контекст, который совместим с контекстом, принятым сопоставителем 350, которые также указывают триггерную фразу, которая соответствует потенциальной  
30 транскрипции голосового ввода.

[0091] В примере, показанном на фиг. 3, сопоставитель 350 принимает контекст, указывающий, что пользовательское устройство 304 имеет медиапроигрыватель, работающий на переднем плане и в режиме аудиопроигрывателя. Сопоставитель 350 также принимает две транскрипции аудиоданных, а именно транскрипцию «play next»

35 и «planets». Сопоставитель 350 идентифицирует потенциальные намерения для инициирования на основе этой информации путем сравнения принятого контекста и транскрипции с контекстом и триггерными фразами, связанными с намерениями, сохраненными в базе 310 данных голосовых действий приложений и базе 320 данных голосовых действий OS. На основе сравнения, сопоставитель 350 идентифицирует два  
40 потенциальных намерения, в том числе намерение для голосового действия «play next», которое вызывает переход приложения медиапроигрывателя к следующей песне, и намерение для голосового действия «planets», которое вызывает открытие приложения, называемого «Planets». Сопоставитель 350 может предоставить информацию блоку 360 создания триггера действия, который идентифицирует два потенциальных намерения,  
45 или сопоставитель 350 может предоставить копию потенциальных намерений блоку 360 создания триггера действия.

[0092] Блок 360 создания триггера действия принимает информацию, которая указывает одно или более потенциальных намерений инициирования в ответ на

голосовой ввод и контекстную информацию, принятые от пользовательского устройства 304. Блок 360 создания триггера действия генерирует триггеры действия для одного или более конкретных намерений, при этом каждый триггер действия содержит данные, которые имеют возможность управлять пользовательским устройством 304 так, чтобы

- 5 выполнить голосовое действие, связанное с соответствующим намерением. В некоторых реализациях блок 360 создания триггера действия генерирует триггер действия для каждого потенциального намерения, идентифицированного сопоставителем 350. В других реализациях блок 360 создания триггера действия может получить указание конкретного намерения, которое должно быть активировано, среди потенциальных 10 намерений и может сгенерировать триггер действия только для этого конкретного намерения. В любой реализации блок 360 создания триггера действия может генерировать триггер действия на основе информации, принятой от сопоставителя 350, и информации, к которой получен доступ в базе 370 данных для хранения данных приложения.

- 15 [0093] Например, блок 360 создания триггера действия может принять намерение для потенциального голосового действия, и блок 360 создания триггера действия может получить доступ к информации в базе 370 данных для хранения данных приложения, что может вызвать выполнение потенциального голосового действия в пользовательском устройстве 304. Например, информация, к которой получен доступ в базе 370 данных 20 для хранения данных приложения, может включать в себя информацию для выполнения или активации активности или действий, связанных с голосовым действием, такую как информация о пользовательском устройстве 304, информация о приложении, связанном с голосовым действием, или информация об операционной системе, запущенной на пользовательское устройство 304. Блок 360 создания триггера действия может 25 использовать информацию, к которой получен доступ, в качестве триггера действия для потенциального голосового действия, или может использовать информацию, к которой получен доступ, для генерации данных триггера действия, которые имеют возможность активации действий или активностей, связанных с потенциальным голосовым действием в пользовательском устройстве 304. В некоторых реализациях 30 системы 300 голосовых действий может не включать в себя блок 360 создания триггера действия или базу 370 данных для хранения данных приложения, так что пользовательскому устройству 304 может быть передано само намерение, и информация в намерении может вызвать выполнение пользовательским устройством 304 действия или активности, связанной с намерением.

- 35 [0094] В примере, показанном на фиг. 3, блок 360 создания триггера действия может принять намерение для голосового действия «play next» для приложения медиапроигрывателя, и намерение для голосового действия «planets» для открытия приложения «Planets». Блок 360 создания триггера действия может сгенерировать триггеры действия для каждого из голосовых действий «play next» и «Planets». Для 40 генерации триггера действия для голосового действия «play next» блок 360 создания триггера действия осуществляет доступ к информации в базе 370 данных для хранения данных приложения, которая относится к приложению медиапроигрывателя, установленному на пользовательском устройстве 304, может получить доступ к информации, которая относится к пользовательскому устройству 304 или конкретной 45 модели пользовательского устройства 304, или может получить доступ к информации, которая относится к операционной системе, которая запущена на пользовательском устройстве 304. Блок 360 создания триггера действия использует принятное намерение и информацию, к которой получен доступ в базе 370 данных для хранения данных

приложения, для генерации триггера действия для голосового действия «play next». Аналогично блок 360 создания триггера действия может генерировать триггер действия для голосового действия «planets» путем осуществления доступа к информации, относящейся к приложению «Planets», пользовательскому устройству 304 или

5 операционной системе, запущенной на пользовательское устройство 304, и может генерировать триггер действия для голосового действия «planets» на основе принятой информации и информации, к которой получен доступ.

[0095] Как обсуждалось выше, в некоторых примерах сопоставитель 350 может идентифицировать более чем одно потенциальное намерение, которое должно быть 10 активировано, и в этих случаях может быть необходимо выбрать конкретное намерение, которое должно быть активировано, среди кандидатов. Механизм 380 устранения неоднозначности может выполнять операции по выбору или приему пользовательского ввода, осуществляющего выбор среди потенциальных намерений. Например, механизм 15 устранения неоднозначности может вывести запрос на пользовательское устройство 304, который просит пользователя 304 выбрать конкретное голосовое действие среди потенциальных голосовых действий, связанных с потенциальными намерениями. В ответ на запрос пользователь 305 может предоставить ввод для пользовательского устройства 304, чтобы выбрать конкретное потенциальное голосовое действие, которое должно быть выполнено. Информация, указывающая выбор пользователя, может быть 20 принята механизмом 380 устранения неоднозначности.

[0096] Альтернативно, механизм 380 устранения неоднозначности может иметь доступ к истории 390 активности пользователя, которая включает в себя информацию о предыдущих голосовых действиях, представленных пользовательскому устройству 304, или других действиях, выполняемых на пользовательском устройстве 304. Например, 25 история 390 активности пользователя может указывать, сколько раз пользователь 305 активировал конкретное потенциальное голосовое действие, сколько раз пользователь 305 выполнял конкретные действия на пользовательском устройстве 304 с помощью или без голосового действия, или может включать в себя другую информацию об использовании пользовательского устройства 304 пользователем 305. В некоторых 30 случаях история 390 активности пользователя может включать в себя информацию, которая указывает историю использования для пользователей помимо пользователя 305. Например, история 390 активности пользователя может включать в себя информацию, которая указывает шаблоны использования всех пользователей, у которых имеется приложение медиапроигрывателя и/или приложение «Planets», установленное 35 на пользовательском устройстве.

[0097] В некоторых реализациях информация в истории 390 активности пользователя может использоваться системой 300 голосовых действий для осуществления выбора среди потенциальных намерений, не запрашивая ввод от пользователя 305. Например, механизм 380 устранения неоднозначности может осуществить доступ к информации 40 в истории 390 активности пользователя, которая указывает, что пользователь 305 ранее выбрал конкретное потенциальное голосовое действие при обеспечении аналогичного голосового ввода, и на основе этой информации механизм 380 устранения неоднозначности может выбрать конкретное потенциальное намерение для активации, не запрашивая ввод от пользователя 305. Аналогично, механизм 380 устранения 45 неоднозначности может установить, что некоторая доля или число пользователей, которые обеспечили голосовой ввод, аналогичный голосовому вводу, принятому от пользователя 305, выбрали для активации конкретное голосовое действие, и механизм 380 устранения неоднозначности может выбрать конкретное потенциальное голосовое

намерение для активации на основе этой информации, не запрашивая ввод от пользователя 305. В некоторых реализациях каждое из потенциальных намерений может иметь соответствующий показатель, который указывает силу взаимосвязи между конкретной триггерной фазой и конкретным потенциальным намерением. Конкретное 5 потенциальное намерение может быть выбрано на основе показателей, присвоенных потенциальным намерениям, которые указывают силу взаимосвязи между потенциальным намерением и принятой триггерной фазой.

[0098] Другие факторы могут учитываться или могут быть включены в историю 390 активности пользователя, которая используется при выборе потенциального намерения.

10 Например, система 300 голосовых действий может отслеживать поведение в пользовательском устройстве 304 после того, как система голосовых действий активирует некоторое потенциальное намерение, чтобы определить, соответствует ли активированное намерение ожиданию пользователя 305. Например, если пользователь 305 выполняет операции для того, чтобы выполнить другое действие в пользовательском 15 устройстве 304 после того, как система 300 голосовых действий активировала некоторое голосовое действие, то это может указывать, что система 300 голосовых действий активировала неправильное потенциальное намерение, и, таким образом, система 300 голосовых действий может добавить информацию в историю 390 активности пользователя, указывающую, что принятый голосовой ввод не должен активировать 20 или должен в меньшей вероятности активировать активированное намерение. В других примерах история 390 активности пользователя может включать в себя информацию, которая указывает, как часто пользователь 305 или другие пользователи выбирают конкретное потенциальное намерение для данного статуса их пользовательского устройства, например, для конкретного статуса, указанного в контекстной информации, 25 предоставленной системе 300 голосовых действий. К другой информации, включенной в историю 390 активности пользователя, может осуществить доступ механизм 380 устранения неоднозначности, и она может использоваться для выбора конкретного потенциального намерения из набора потенциальных намерений.

[0099] Для примера, показанного на фиг. 3, на основе установления, что

30 потенциальные намерения для голосового ввода, обеспеченного пользователем 305, включают в себя намерение для голосового действия «play next» для приложения медиапроигрывателя или намерение для голосового действия «planets» для открытия приложения «Planets», механизм 380 устранения неоднозначности может предоставить информацию пользовательскому устройству 305, которая просит пользователя выбрать 35 конкретное потенциальное голосовое действие. Как показано, запрос может быть выведен на дисплее пользовательского устройства 304, который просит пользователя 305 выбрать либо голосовое действие «Следующая песня», либо голосовое действие «Открыть 'Planets'». Пользователь 305 обеспечивает ввод на пользовательском устройстве 304 для выбора опции «Следующая песня». Данные, указывающие выбор 40 пользователя, передаются механизму 380 устранения неоднозначности, и при приеме данных, указывающих выбор пользователя, механизм 380 устранения неоднозначности определяет, что необходимо активировать намерение голосового действия «play next». В некоторых примерах механизм 380 устранения неоднозначности также может хранить информацию в истории 390 активности пользователя, указывающую, что пользователь 45 305 выбрал активацию намерения для голосового действия «play next», а не намерение для голосового действия «planets». Механизм 380 устранения неоднозначности может полагаться на информацию, хранящуюся в истории 390 активности пользователя, для автоматического определения потенциального голосового действия, которое должно

быть активировано, если аналогичный голосовой ввод позже принят системой 300 голосовых действий от пользовательского устройства 304.

[00100] В альтернативной реализации примера, показанного на фиг. 3, механизм 380 устранения неоднозначности может выбрать конкретное потенциальное намерение для активации на основе осуществления доступа к истории 390 активности пользователя. Например, история 390 активности пользователя может указывать, что пользователь 305 пользовательского устройства 304 ранее обеспечил аналогичный голосовой ввод и выбрал активацию намерения для голосового действия «Следующая песня». На основе истории 390 активности пользователя, указывающей, что ранее было активировано намерение для голосового действия «play next» после приема аналогичного голосового ввода, механизм 380 устранения неоднозначности может определить, что необходимо активировать намерение для голосового действия «play next», не запрашивая дополнительный ввод от пользователя 305.

[00101] После того, как конкретное потенциальное намерение было выбрано и, в случае необходимости, триггер действия для выбранного потенциального голосового действия сгенерирован, система 300 голосовых действий передает информацию пользовательскому устройству 304, что вызывает выполнение выбранного голосового действия. Например, механизм 380 устранения неоднозначности или другой компонент системы 300 голосовых действий может передать намерение или данные триггера действия для выбранного голосового действия пользовательскому устройству 304, чтобы вызвать выполнение пользовательским устройством 304 выбранного голосового действия.

[00102] Например, в примере, показанном на фиг. 3, после того, как пользователь 305 выбрал опцию «Следующая песня» для выбора голосового действия «play next», что вызывает переход приложения медиапроигрывателя к следующей песне, и после того, как был сгенерирован триггер действия, который включает в себя данные для управления приложением медиапроигрывателя для перехода к следующей песне, механизм 380 устранения неоднозначности может передать триггер действия пользовательскому устройству 304. Пользовательское устройство 304 может принять триггер действия, и триггер действия может вызвать выполнение пользовательским устройством 304 голосового действия «play next» путем обеспечения перехода приложения медиапроигрывателя к следующей песне.

[00103] В некоторых реализациях уведомление, указывающее активированное голосовое действие, может быть обеспечено для вывода на пользовательском устройстве 304. Например, система 300 голосовых действий или OS пользовательского устройства 304 может вызвать вывод уведомления на пользовательском устройстве 304, которое идентифицирует голосовое действие, которое было активировано. В некоторых случаях уведомление, идентифицирующее активированное голосовое действие, может быть обеспечено таким образом, что пользователю 305 очевидно, что приложение, связанное с активированным голосовым действием, не отвечает за детектирование голосового действия или активацию голосового действия. Например, уведомление, обеспеченное на пользовательском устройстве 304, может использовать тон голоса, который распознается как связанный с системой 300 голосовых действий, пользовательским устройством 304 или OS пользовательского устройства 304, а не приложением медиапроигрывателя, работающим на пользовательском устройстве 304. В другом примере уведомление может быть представлено таким способом, который отличает систему или приложение, обрабатывающее голосовой ввод от пользователя 305, от приложения, которое выполняет активированное голосовое действие. Например,

уведомление может заявлять: «Вы активировали голосовое действие 'play next'. Теперь я сделаю так, что приложение медиапроигрывателя перейдет к следующей песне», тем самым указывая пользователю 305, что приложение медиапроигрывателя не отвечало за обработку голосового ввода.

5 [00104] Фиг. 4 изображает блок-схему последовательности операций для иллюстративного процесса 400, выполняемого системой голосовых действий. В некоторых примерах процесс 400 на фиг. 4 может выполняться системой 100 голосовых действий на фиг. 1 или компонентами системы 100 голосовых действий на фиг. 1. Процесс 400 описывает процесс с помощью которого разработчик приложений может назначить 10 новое голосовое действие для приложения, которое было установлено на одном или более пользовательских устройствах, и с помощью которого пользователь пользовательского устройства, имеющий установленное приложение, может активировать новое голосовое действие для выполнения операций, связанных с новым голосовым действием, приложением на пользовательском устройстве.

15 [00105] Вначале системой голосовых действий принимаются данные, которые задают новое голосовое действие для программного приложения, которое отличается от системы голосовых действий, при этом данные содержат одну или более операций для выполнения нового голосового действия и один или более триггерных элементов для активации нового голосового действия (402). Например, разработчик 101a-101n 20 приложений, связанный с программным приложением, может представить данные, задающие новое голосовое действие, системе 100 голосовых действий с использованием терминала 102a-102n.

[00106] Данные, задающие новое голосовое действие, могут включать в себя данные, указывающие одну или более операций для выполнения нового голосового действия 25 и один или более триггерных элементов для активации нового голосового действия. Например, данные, задающие новое голосовое действие, могут указывать триггерный элемент «play next» и может указывать, что новое голосовое действие должно выполнять операции, которые вызывают переход приложения медиапроигрывателя к следующей песне. Данные также могут указывать дополнительную информацию, относящуюся к 30 новому голосовому действию, такую как приложение, к которому относится новое голосовое действие, контекст, когда новое голосовое действие должно быть разрешено для использования пользователем, или другую информацию. Например, данные, задающие голосовое действие «play next», могут также указывать, что голосовое действие относится к приложению медиапроигрывателя, и могут указывать, что голосовое 35 действие должно быть разрешено, когда приложение медиапроигрывателя работает на переднем плане и находится в режиме аудиопроигрывателя.

[00107] Контекстное намерение голосового действия для программного приложения генерируется системой голосовых действий на основе по меньшей мере принятых данных, при этом контекстное намерение голосового действия содержит данные, 40 которые при приеме программным приложением запрашивают выполнение программным приложением одной или более операций нового голосового действия (404). Например, система 100 голосовых действий на фиг. 1 может принять данные, указывающие новое голосовое действие, и может сгенерировать контекстное намерение голосового действия для нового голосового действия на основе принятых данных. 45 Система 100 голосовых действий может сохранить контекстное намерение голосового действия в базе 110 данных голосовых действий приложений. В некоторых реализациях система 100 голосовых действий может сгенерировать намерение с использованием процесса, аналогичного описанному в отношении фиг. 2, путем проверки принятых

данных и введения принятой грамматики для генерации контекстного намерения голосового действия. В некоторых реализациях сгенерированное контекстное намерение голосового действия может хранить информацию, которая может управлять программным приложением так, чтобы выполнить одну или более операций для выполнения нового голосового действия.

5 [00108] Контекстное намерение голосового действия связывается с одним или более триггерными элементами для нового голосового действия системой голосовых действий (406). Например, после создания контекстного намерения голосового действия для нового голосового действия системой 100 голосовых действий, система 100 голосовых  
10 действий может связать контекстное намерение голосового действия для нового голосового действия с одним или более триггерными элементами, указанными в принятых данных. Контекстное намерение голосового действия, имеющее связанный один или более триггерных элементов, может быть сохранено в базе 110 данных голосовых действий приложений. В некоторых реализациях, связывание одного или  
15 более триггерных элементов с контекстным намерением голосового действия может включать в себя указание одного или более триггерных элементов в контекстном намерении голосового действия. В других реализациях связывание может быть произведено иным связыванием одного или более триггерных элементов с контекстным намерением голосового действия, так что получение одного из одного или более  
20 триггерных элементов приводит к идентификации контекстного намерения голосового действия.

[00109] Система голосовых действий выполнена с возможностью приема указания относительно высказывания пользователя, полученного устройством с установленным программным приложением (408). Например, система 100 голосовых действий может  
25 быть выполнена с возможностью приема указания, что от пользовательского устройства 104 было принято высказывание, при этом высказывание получено пользовательским устройством 104 в виде голосового ввода от пользователя 105. Пользовательское устройство 104 может иметь уже установленное программное приложение, связанное с новым голосовым действием, например, приложение медиапроигрывателя,  
30 обсуждавшееся в отношении фиг. 1.

[00110] Система голосовых действий выполнена с возможностью установления, что транскрипция высказывания пользователя соответствует одному или более триггерным элементам, связанным с контекстным намерением голосового действия (410). Например, система 100 голосовых действий может быть выполнена с возможностью приема  
35 аудиоданных, соответствующих высказыванию пользователя, и может генерировать транскрипцию высказывания пользователя с использованием механизма 130 распознавания речи системы 100 голосовых действий. Транскрипция высказывания пользователя, сгенерированная механизмом 130 распознавания речи, может сравниваться с триггерными элементами, связанными с множеством контекстных  
40 намерений голосовых действий, в том числе контекстным намерением голосового действия для нового голосового действия и контекстными намерениями голосовых действий для других голосовых действий.

[00111] В некоторых случаях сравнение может выполняться селектором 150 голосового действия на фиг. 1, при этом селектор 150 голосового действия может  
45 сравнивать транскрипцию высказывания пользователя с триггерными элементами для контекстных намерений голосового действия, сохраненных в базе 110 данных голосовых действий приложений или базе 130 данных голосовых действий OS. На основе сравнения селектор 150 голосового действия может установить, что транскрипция высказывания

пользователя соответствует одному или более из одного или более триггерных элементов, связанных с контекстным намерением голосового действия нового голосового действия. Например, механизм 130 распознавания речи может сгенерировать транскрипцию для высказывания пользователя, принятого от пользовательского устройства 104, которая имеет вид «play next». Селектор 150 голосового действия может сравнить голосовое действие «play next» с триггерными элементами, связанными с потенциальными контекстными намерениями голосовых действий для установления, что транскрипция соответствует триггерной фразе «play next», связанной с контекстным намерением голосового действия для нового голосового действия «play next».

- 10 [00112] Система голосовых действий выполнена с возможностью обеспечения контекстного намерения голосового действия для конкретного устройства, тем самым запроса, чтобы программное приложение, установленное на конкретном устройстве, выполнило одну или более операций нового голосового действия (412). Например, в ответ на определение, что транскрипция высказывания пользователя, сгенерированная 15 механизмом 130 распознавания речи системы 100 голосовых действий, соответствует одному или более триггерным элементам, связанным с контекстным намерением голосового действия для нового голосового действия, селектор 150 голосового действия может обеспечить контекстное намерение голосового действия для нового голосового действия пользовательскому устройству 104. Контекстное намерение голосового 20 действия для нового голосового действия может эффективно запросить, чтобы программное приложение, установленное на пользовательском устройстве 104, выполнило одну или более операций нового голосового действия. Например, контекстное намерение голосового действия может включать в себя данные, которые активируют одну или более операций, связанных с новым голосовым действием, при 25 приеме пользовательским устройством 104 контекстного намерения голосового действия.

[00113] Выше был описан ряд реализаций. Однако следует понимать, что могут быть сделаны различные модификации, не отступая от сущности и объема раскрытия. Например, могут использоваться различные формы последовательности операций, 30 показанной выше, с измененным порядком этапов, добавленными или убранными этапами. Соответственно, другие реализации находятся в пределах объема прилагаемой формулы изобретения.

[00114] Для случаев, в которых системы и/или способы, обсуждавшиеся здесь, могут собирать персональные данные о пользователях или могут использовать персональные 35 данные, пользователям может быть предоставлена возможность управлять, собирают ли программы или функциональные возможности персональные данные, например, информацию о социальной сети пользователя, социальных действиях или активностях, профессии, предпочтениях или текущем местоположении, или управлять, может ли и/ или как система и/или способы могут выполнять операции, более релевантные 40 пользователю. Кроме того, некоторые данные могут обезличиваться одним или более способами перед сохранением или использованием, так что информация с возможностью персональной идентификации убирается. Например, идентификационные данные 45 пользователя могут обезличиваться так, чтобы никакая информация с возможностью персональной идентификации не могла быть определена для пользователя, или географическое местоположение пользователя может обобщаться, когда получается информация о местоположении, например, до города, почтового индекса или уровня региона, так что конкретное местоположение пользователя не может быть определено. Таким образом, пользователь может управлять тем, как информация о нем или ней

собирается и используется.

[00115] Хотя приведенные выше варианты осуществления преимущественно были описаны со ссылкой на разработку голосовых действий для использования с приложениями, установленными на пользовательских устройствах, описанные признаки 5 также могут использоваться в отношении машин, других устройств, роботов или других систем. Например, система 100 голосовых действий на фиг. 1 может использоваться для разработки и реализации голосовых действий для взаимодействия с машинами, где машины имеет связанную вычислительную систему, может использоваться для разработки и реализации голосовых действий для взаимодействия с роботом или 10 системой, имеющей роботизированные компоненты, может использоваться для разработки и реализации голосовых действий для взаимодействия с приборами, развлекательными системами или другими устройствами, или может использоваться для разработки и реализации голосовых действий для взаимодействия с транспортным средством или другой системой транспортировки.

[00116] Варианты осуществления и все функциональные операции, описанные в этом описании, могут быть реализованы в цифровых электронных схемах, или в программном обеспечении, микропрограммном обеспечении, или аппаратном обеспечении, в том числе структурах, раскрытых в этом описании и его структурных эквивалентах, или в комбинациях одного или более из этого. Варианты осуществления могут быть 15 реализованы как один или более компьютерных программных продуктов, то есть один или более модулей инструкций компьютерных программ, закодированных на машиночитаемой среде для исполнения устройством обработки данных или управления работой устройства обработки данных. Машиночитаемая среда может быть машиночитаемым устройством хранения, машиночитаемой запоминающей подложкой, 20 запоминающим устройством, составом вещества, воздействующей на машиночитаемый распространяющийся сигнал, или комбинацией одного или более из них. Термин «устройство обработки данных» охватывает все аппараты, устройства и машины для обработки данных, в том числе, в качестве примера, программируемый процессор, компьютер или множество процессоров или компьютеров. Устройство может включать 25 в себя, в дополнение к аппаратному обеспечению, код, который создает среду исполнения для рассматриваемой компьютерной программы, например, код, который составляет микропрограммное обеспечение процессора, стек протоколов, систему управления базами данных, операционную систему или комбинацию одного или более из них. Распространяющийся сигнал является искусственно сгенерированным сигналом, 30 например, машинно-генерируемый электрический, оптический или электромагнитный сигнал, который генерируется для кодирования информации для передачи подходящему приемному устройству.

[00117] Компьютерная программа (также известная как программа, программное обеспечение, программное приложение, сценарий или код) может быть записана на 35 любом языке программирования, в том числе компилируемых или интерпретируемых языках, и она может развертываться в любой форме, в том числе как автономная программа или как модуль, компонент, подпрограмма или другой блок, пригодный для использования в вычислительной среде. Компьютерная программа не обязательно соответствует файлу в файловой системе. Программа может храниться в части файла, 40 который содержит другие программы или данные (например, один или более сценариев, сохраненных в документе на языке разметки), в одном файле, предназначенном для рассматриваемой программы, или в множестве согласованных файлов (например, файлах, которые хранят один или более модулей, подпрограмм или частей кода).

Компьютерная программа может быть развернута для исполнения на одном компьютере или на множестве компьютеров, которые расположены в одном месте или распределены между несколькими местами и соединены с помощью сети связи.

[00118] Последовательность процессов и логический поток, описанные в этом

5 описании, могут выполняться одним или более программируемыми процессорами, исполняющими одну или более компьютерных программ для выполнения функций путем работы с входными данными и генерации выходных данных. Последовательность процессов и логический поток также могут выполняться специализированной схемой логики, и устройство может также быть реализовано как специализированная схема 10 логики, например, FPGA (программируемая пользователем вентильная матрица) или ASIC (специализированная интегральная схема).

[00119] Процессоры, подходящие для исполнения компьютерной программы,

включают, в качестве примера, как универсальные и специализированные 15 микропроцессоры, так и любой один или более процессоров любого типа цифрового компьютера. Как правило, процессор будет принимать инструкции и данные из памяти только для чтения или оперативной памяти, или их обоих.

[00120] Важными элементами компьютера являются процессор для выполнения инструкций и одно или более запоминающих устройств для хранения инструкций и 20 данных. Как правило, компьютер также будет включать в себя или будет функционально соединен для приема данных или передачи данных, или и то, и другое, одному или более устройствам хранения большой емкости для хранения данных, например, магнитные, 25 магнитооптические диски или оптические диски. Однако компьютер не обязательно должен иметь такие устройства. Кроме того, компьютер может быть встроен в другое устройство, например, планшетный компьютер, мобильный телефон, персональный цифровой помощник (PDA), мобильный аудиопроигрыватель, приемник глобальной 30 навигационной системы (GPS), если перечислять лишь некоторые из них.

Машиночитаемые среды, подходящие для хранения инструкций компьютерной программы и данных, включают в себя все формы энергонезависимой памяти, 35 устройства хранения данных и запоминающие устройства, в том числе в качестве

примера полупроводниковые запоминающие устройства, например, EPROM, EEPROM и устройства флэш-памяти; магнитные диски, например, внутренние жесткие диски или 40 съемные диски; магнитооптические диски; и диски CD-ROM и DVD-ROM. Процессор и память могут дополняться или могут быть встроены в специализированные логические схемы.

[00121] Для обеспечения взаимодействия с пользователем варианты осуществления 45 могут быть реализованы на компьютере, имеющем устройство отображения, например, CRT (электронно-лучевую трубку) или ЖК-дисплей (жидкокристаллический дисплей) для отображения информации пользователю, а также клавиатуру и указательное устройство, например, мышь или шаровой манипулятор, с помощью которого

пользователь может обеспечить ввод для компьютера. Также могут использовать 50 другие типы устройств для обеспечения взаимодействия с пользователем; например, обратная связь, предоставляемая пользователю, может быть любой формой сенсорной обратной связи, например, визуальной обратной связью, акустической обратной связью или осензительной обратной связью; и ввод от пользователя может приниматься в любой 55 форме, в том числе акустический, голосовой или тактильный ввод.

[00122] Варианты осуществления могут быть реализованы в вычислительной системе, которая включает в себя серверный компонент, например, сервер данных, или включает в себя компонент промежуточного программного обеспечения, например, сервер

приложений, или включает в себя интерфейсный компонент, например, клиентский компьютер, имеющий графический пользовательский интерфейс, или веб-браузер, с помощью которого пользователь может взаимодействовать с реализацией, или любую комбинацию одного или более из такого серверного компонента, компонента

5 промежуточного программного обеспечения или интерфейсного компонента. Компоненты системы могут быть соединены между собой с помощью любой формы или среды передачи цифровых данных, например, сети связи. Примеры сетей связи включают в себя локальную сеть («LAN») и глобальную сеть («WAN»), например, Интернет.

10 [00123] Вычислительная система может включать в себя клиенты и серверы. Клиент и сервер в общем случае удалены друг от друга и, как правило, взаимодействует через сеть связи. Взаимосвязь клиента и сервера возникает посредством компьютерных программ, выполняющихся на соответствующих компьютерах и имеющих взаимосвязь клиент-сервер друг с другом.

15 [00124] Хотя это описание содержит много конкретных подробностей, они не должны рассматриваться как ограничение объема раскрытия или формулы изобретения, а должны рассматриваться скорее как описание признаков, характерных для конкретных вариантов осуществления. Некоторые признаки, которые описаны в этом описании в контексте отдельных вариантов осуществления, также могут быть реализованы в 20 комбинации в одном варианте осуществления. И наоборот, различные признаки, которые описаны в контексте одного варианта осуществления, также могут быть реализованы во множестве вариантов осуществления отдельно или в любой подходящей подкомбинации. Кроме того, хотя признаки выше могли быть описаны как действующие 25 в некоторых комбинациях и даже первоначально заявлены как таковые, один или более признаков из заявленной комбинации в некоторых случаях могут быть исключены из комбинации, и заявленная комбинация может быть обращена в подкомбинацию или вариант подкомбинации.

[00125] Аналогично, хотя операции изображены на чертежах в конкретном порядке, это не должно рассматриваться как требование, чтобы такие операции выполнялись в 30 конкретном показанном порядке или в последовательном порядке, или чтобы все изображенные операции выполнялись для достижения требуемых результатов. В некоторых обстоятельствах может быть выгодна многозадачная и параллельная обработка. Кроме того, разделение различных системных компонентов в вариантах осуществления, описанных выше, не должно рассматриваться как требование иметь 35 такое разделение во всех вариантах осуществления, и следует понимать, что описанные компоненты программы и системы, в общем случае, могут быть интегрированы вместе в одном программном продукте или собраны во множество программных продуктов.

[00126] В каждом случае, когда упоминался HTML-файл, могут быть подставлены 40 другие типы файлов или форматы. Например, HTML-файл может быть заменен XML, JSON, простым текстом или другими типами файлов. Кроме того, там, где упоминалась таблица или хэш-таблица, могут использоваться другие структуры данных (такие как электронные таблицы, реляционные базы данных или структурированные файлы).

[00127] Дополнительные реализации кратко описаны в следующих примерах:

[00128] Пример 1: реализованные с помощью компьютера способ, содержащий 45 прием системой голосовых действий данных, указывающих новое голосовое действие для программного приложения, отличающегося от упомянутой системы голосовых действий, данных, содержащих одну или более операций для выполнения нового голосового действия и один или более триггерных элементов для активации нового

голосового действия;

для программного приложения на основе по меньшей мере принятых данных, при этом контекстное намерение голосового действия содержит данные, которые при приеме 5 программным приложением запрашиваются, чтобы программное приложение выполнило одну или более операций нового голосового действия;

связывание системой голосовых действий контекстного намерения голосового действия с одним или более триггерными элементами для нового голосового действия; при этом система голосовых действий выполнена с возможностью:

10 приема указания относительно высказывания пользователя, полученного устройством, на котором установлено программное приложение,

определения, что транскрипция высказывания пользователя соответствует одному или более триггерным элементам, связанным с контекстным намерением голосового действия; и

15 в ответ на это определение, предоставления устройству контекстного намерения голосового действия, тем самым запроса, чтобы программное приложение, установленное на устройстве, выполнило одну или более операций нового голосового действия.

[00129] Пример 2: Компьютерно-реализуемый способ согласно примеру 1, в котором

20 принятые данные указывают контекст, этот контекст указывает статус устройства или программного приложения, когда новое голосовое действие разрешено.

[00130] Пример 3: Компьютерно-реализуемый способ согласно примеру 2, в котором контекст указывает, что программное приложение работает на переднем плане устройства, на котором установлено программное приложение.

25 [00131] Пример 4: Компьютерно-реализуемый способ согласно примеру 2, в котором контекст указывает, что программное приложение работает в фоновом режиме устройства, на котором установлено программное приложение.

[00132] Пример 5: Компьютерно-реализуемый способ согласно примеру 2, в котором контекст указывает, что программное приложение выполняет конкретную активность.

30 [00133] Пример 6: Компьютерно-реализуемый способ согласно примеру 2, в котором контекст указывает, что конкретная активность, которую выполняет программное приложение, находится в конкретном состоянии активности.

[00134] Пример 7: Компьютерно-реализуемый способ согласно одному из примеров 2-6, содержащий:

35 связывание контекстного намерения голосового действия с контекстом для нового голосового действия;

прием системой голосовых действий контекстной информации, указывающей статус конкретного устройства, имеющего установленное программное приложение, или программного приложения, установленного на конкретном устройстве;

40 определение, что контекстная информация удовлетворяет контексту для нового голосового действия; и

в ответ на определение, что транскрипция высказывания пользователя, полученная конкретным устройством, соответствует одному или более триггерным элементам, связанным с контекстным намерением голосового действия, и что контекстная

45 информация удовлетворяет контексту, связанному с контекстным намерением голосового действия, предоставление системой голосовых действий конкретному устройству контекстного намерения голосового действия, тем самым запрос, чтобы программное приложение, установленное на конкретном устройстве, выполнило одну

или более операций нового голосового действия.

[00135] Пример 8: Компьютерно-реализуемый способ согласно примеру 7, в котором прием контекстной информации, указывающей статус конкретного устройства или программного приложения, установленного на конкретном устройстве, содержит:

5 предоставление системой голосовых действий конкретному устройству запроса на конкретную контекстную информацию; и

прием конкретной контекстной информации в ответ на запрос.

[00136] Пример 9: Компьютерно-реализуемый способ согласно одному из примеров 7 или 8, содержащий:

10 определение, что контекстная информация удовлетворяет контексту для второго голосового действия, и что транскрипция высказывания пользователя соответствует одному или более триггерным элементам, которые связаны с намерением для второго голосового действия, при этом намерение для второго голосового действия указывает одну или более операций для выполнения второго голосового действия;

15 в ответ на это определение, выбор голосового действия между новым голосовым действием и вторым голосовым действием; и

предоставление системой голосовых действий конкретному устройству намерения, связанного с выбранным голосовым действием, тем самым запрос, чтобы программное приложение, установленное на конкретном устройстве, выполнило одну или более

20 операций выбранного голосового действия.

[00137] Пример 10: Компьютерно-реализуемый способ согласно примеру 9, в котором выбор выбранного голосового действия между новым голосовым действием и вторым голосовым действием содержит выбор выбранного голосового действия в ответ на прием данных, указывающих выбор пользователем одного из: нового голосового

25 действия или второго голосового действия.

[00138] Пример 11: Компьютерно-реализуемый способ согласно примеру 9 или 10, в котором выбор выбранного голосового действие между новым голосовым действием и вторым голосовым действием содержит:

30 присваивание показателя каждому из: новому голосовому действию и второму голосовому действию; и

выбор выбранного голосового действия на основе по меньшей мере показателя, присвоенного каждому из: новому голосовому действию и второму голосовому действию.

[00139] Пример 12: Компьютерно-реализуемый способ согласно одному из примеров 35 9-11, в котором выбор выбранного голосового действия между новым голосовым действием и вторым голосовым действием содержит выбор выбранного голосового действия в ответ на определение, что программное приложение, связанное с выбранным голосовым действием, работает на переднем плане.

[00140] Пример 13: Компьютерно-реализуемый способ согласно одному из примеров 40 1-12, в котором генерация контекстного намерения голосового действия для программного приложения содержит определение, что одна или более операций нового голосового действия могут быть выполнены программным приложением.

[00141] Пример 14: Компьютерно-реализуемый способ согласно одному из примеров 1-13, содержащий:

45 определение, что транскрипция высказывания пользователя, полученная конкретным устройством, имеющим установленное программное приложение, аналогична одному или более триггерным элементам, связанным с контекстным намерением голосового действия;

в ответ на это определение, предоставление системой голосовых действий конкретному устройству данных, указывающих запрос на пользовательский ввод, который подтверждает, указывало ли высказывание пользователя один или более триггерных элементов или должно было вызвать выполнение программным

5 приложением нового голосового действия;

в ответ на этот запрос, прием системой голосовых действий и от конкретного устройства данных, указывающих подтверждение; и

в ответ на прием данных, указывающих подтверждение, предоставление системой голосовых действий конкретному устройству контекстного намерения голосового

10 действия, тем самым запроса, чтобы программное приложение, установленное на конкретном устройстве, выполнило одну или более операций нового голосового действия.

[00142] Пример 15: Компьютерно-реализуемый способ согласно одному из примеров 1-14, содержащий:

15 прием системой голосовых действий от разработчика, связанного с программным приложением, запроса на развертывание нового голосового действия для программного приложения; и

в ответ на этот запрос, развертывание нового голосового действия для программного приложения в ответ на этот запрос, при этом развертывание нового голосового действия

20 для программного приложения разрешает новое голосовое действие для программного приложения.

[00143] Пример 16: Компьютерно-реализуемый способ согласно одному из примеров 1-15, содержащий:

прием системой голосовых действий от разработчика, связанного с программным

25 приложением, запроса на отмену развертывания нового голосового действия для программного приложения; и

в ответ на этот запрос, отмену развертывания нового голосового действия для программного приложения в ответ на этот запрос, при этом отмена развертывания нового голосового действия для программного приложения запрещает новое голосовое

30 действие для программного приложения.

[00144] Пример 17: Компьютерно-реализуемый способ согласно одному из примеров 1-16, содержащий:

прием системой голосовых действий от разработчика, связанного с программным приложением, запроса на разрешение тестирования нового голосового действия, при

35 этом запрос указывает одно или более устройств, для которых должно быть разрешено новое голосовое действие; и

в ответ на этот запрос, разрешение нового голосового действия для одного или более указанных устройств, при этом новое голосовое действие запрещено для устройств, которые не включены в указанные устройства.

40 [00145] Пример 18: Компьютерно-реализуемый способ согласно одному из примеров 1-17, содержащий:

прием системой голосовых действий указания относительно высказывания пользователя, полученного конкретным устройством, имеющим установленное программное приложение;

45 определение системой голосовых действий, что транскрипция высказывания пользователя соответствует одному или более триггерным элементам, связанным с контекстным намерением голосового действия; и

в ответ на это определение, предоставление системой голосовых действий

конкретному устройству контекстного намерения голосового действия, тем самым запроса, чтобы программное приложение, установленное на конкретном устройстве, выполнило одну или более операций нового голосового действия.

[00146] Пример 19: Энергонезависимое машиночитаемое устройство хранения,

5 хранящее инструкции, исполнимые одним или более компьютерами, которые при таком исполнении вызывают выполнение одним или более компьютерами операций, содержащих этапы примеров 1-18.

[00147] Выше были описаны способы, системы и устройство для приема системой 10 голосовых действий данных, указывающих новое голосовое действие для приложения, отличающегося от системы голосовых действий. Намерение голосового действия для

приложения генерируется на основе по меньшей мере этих данных, при этом намерение голосового действия содержит данные, которые при приеме приложением запрашиваются, чтобы приложение выполнило одну или более операций, указанных для нового голосового действия. Намерение голосового действия связано с триггерными

15 элементами, указанными для нового голосового действия. Система голосовых действий выполнена с возможностью приема указания относительно высказывания пользователя, полученного устройством, имеющим установленное приложение, и она определяет, что транскрипция высказывания пользователя соответствует триггерным элементам, связанным с намерением голосового действия. В ответ на это определение система

20 голосовых действий предоставляет намерение голосового действия устройству.

[00148] Таким образом, были описаны конкретные варианты осуществления. Другие варианты осуществления находятся в пределах объема формулы изобретения. Например, 25 действия, перечисленные в формуле изобретения, могут выполняться в другом порядке и тем не менее обеспечивать желаемые результаты.

25

### (57) Формула изобретения

1. Компьютерно-реализуемый способ назначения новых голосовых действий для 30 программного приложения, которое ранее было установлено на пользовательских устройствах, при этом голосовое действие задает одну или более операций, выполняемых в программном приложении, причем способ выполняется в системе голосовых действий, отличающейся от пользовательских устройств, при этом способ содержит этапы, на которых:

35 принимают посредством системы голосовых действий данные, указывающие новое голосовое действие для программного приложения, отличающегося от упомянутой системы голосовых действий, причем данные содержат (1) одну или более операций для выполнения нового голосового действия, (2) один или более триггерных элементов для активации нового голосового действия и (3) контекст, указывающий статус 40 пользовательского устройства или программного приложения, когда новое голосовое действие разрешено;

45 генерируют посредством системы голосовых действий пассивную структуру данных голосового действия для программного приложения на основе, по меньшей мере, принятых данных, при этом пассивная структура данных голосового действия содержит данные, которые при приеме программным приложением запрашиваются, чтобы программное приложение выполнило одну или более операций нового голосового действия;

связывают посредством системы голосовых действий пассивную структуру данных голосового действия с контекстом и с одним или более триггерными элементами для нового голосового действия;

при этом система голосовых действий выполнена с возможностью:

принимать от пользовательского устройства указание относительно (1) высказывания пользователя, полученного пользовательским устройством, на котором установлено программное приложение, и (2) контекстной информации, указывающей статус 5 пользователя, на котором установлено программное приложение, или программного приложения, установленного на пользовательском устройстве;

определять, что (1) транскрипция высказывания пользователя соответствует одному или более триггерным элементам, связанным с пассивной структурой данных голосового действия, и что (2) контекстная информация удовлетворяет контексту для нового 10 голосового действия; и

в ответ на это определение подавать в пользовательское устройство пассивную структуру данных голосового действия, тем самым запрашивая, чтобы программное приложение, установленное на пользовательском устройстве, выполнило одну или более операций нового голосового действия.

15 2. Компьютерно-реализуемый способ по п. 1, в котором контекст указывает, что программное приложение работает на переднем плане устройства, на котором установлено программное приложение.

3. Компьютерно-реализуемый способ по п. 1, в котором контекст указывает, что 20 программное приложение работает в фоновом режиме устройства, на котором установлено программное приложение.

4. Компьютерно-реализуемый способ по п. 1, в котором контекст указывает, что программное приложение выполняет конкретную активность.

5. Компьютерно-реализуемый способ по п. 1, в котором контекст указывает, что конкретная активность, которую выполняет программное приложение, находится в 25 конкретном состоянии активности.

6. Компьютерно-реализуемый способ по п. 1, содержащий этапы, на которых: в ответ на определение того, что транскрипция высказывания пользователя, полученная конкретным устройством, соответствует одному или более триггерным элементам, связанным с контекстным намерением голосового действия, и что контекстная 30 информация удовлетворяет контексту, связанному с контекстным намерением голосового действия, предоставляют посредством системы голосовых действий конкретному устройству контекстное намерение голосового действия, тем самым запрашивая, чтобы программное приложение, установленное на конкретном устройстве, выполнило одну или более операций нового голосового действия.

35 7. Компьютерно-реализуемый способ по п. 6, в котором упомянутый прием контекстной информации, указывающей статус конкретного устройства или программного приложения, установленного на конкретном устройстве, содержит этапы, на которых:

подают посредством системы голосовых действий в конкретное устройство запрос

40 на конкретную контекстную информацию; и

принимают конкретную контекстную информацию в ответ на данный запрос.

8. Компьютерно-реализуемый способ по п. 6, содержащий этапы, на которых:

45 определяют, что контекстная информация удовлетворяет контексту для второго голосового действия и что транскрипция высказывания пользователя соответствует одному или более триггерным элементам, которые связаны с намерением для второго голосового действия, при этом намерение для второго голосового действия указывает одну или более операций для выполнения второго голосового действия;

в ответ на это определение выбирают голосовое действие между новым голосовым

действием и вторым голосовым действием; и

предоставляют посредством системы голосовых действий конкретному устройству намерение, связанное с выбранным голосовым действием, тем самым запрашивая, чтобы программное приложение, установленное на конкретном устройстве, выполнило

5 одну или более операций выбранного голосового действия.

9. Компьютерно-реализуемый способ по п. 8, в котором упомянутый выбор голосового действия, выбиравшего между новым голосовым действием и вторым голосовым действием, содержит этап, на котором выбирают выбиравшее голосовое действие в ответ на прием данных, указывающих выбор пользователем одного из

10 нового голосового действия и второго голосового действия.

10. Компьютерно-реализуемый способ по п. 8, в котором упомянутый выбор голосового действия, выбиравшего между новым голосовым действием и вторым голосовым действием, содержит этапы, на которых:

присваивают показатель каждому из нового голосового действия и второго

15 голосового действия; и

выбирают выбиравшее голосовое действие на основе, по меньшей мере, показателя, присвоенного каждому из нового голосового действия и второго голосового действия.

11. Компьютерно-реализуемый способ по п. 8, в котором упомянутый выбор голосового действия, выбиравшего между новым голосовым действием и вторым

20 голосовым действием, содержит этап, на котором выбирают выбиравшее голосовое действие в ответ на определение того, что программное приложение, связанное с выбиравшим голосовым действием, работает на переднем плане.

12. Компьютерно-реализуемый способ по п. 1, в котором упомянутое генерирование контекстного намерения голосового действия для программного приложения содержит

25 этап, на котором определяют, что одна или более операций нового голосового действия могут быть выполнены программным приложением.

13. Компьютерно-реализуемый способ по п. 1, содержащий этапы, на которых:

определяют, что транскрипция высказывания пользователя, полученная конкретным устройством, имеющим установленное программное приложение, аналогична одному

30 или более триггерным элементам, связанным с контекстным намерением голосового действия;

в ответ на это определение подают посредством системы голосовых действий в конкретное устройство данные, указывающие запрос на пользовательский ввод, который подтверждает, указывало ли высказывание пользователя один или более триггерных

35 элементов или должно было вызвать выполнение программным приложением нового голосового действия;

в ответ на этот запрос принимают посредством системы голосовых действий и от конкретного устройства данные, указывающие подтверждение; и

в ответ на прием данных, указывающих подтверждение, подают посредством системы

40 голосовых действий в конкретное устройство контекстное намерение голосового действия, тем самым запрашивая, чтобы программное приложение, установленное на конкретном устройстве, выполнило одну или более операций нового голосового действия.

14. Компьютерно-реализуемый способ по п. 1, содержащий этапы, на которых:

45 принимают посредством системы голосовых действий от разработчика, связанного с программным приложением, запрос на развертывание нового голосового действия для программного приложения; и

в ответ на этот запрос развертывают новое голосовое действие для программного

приложения в ответ на данный запрос, при этом развертывание нового голосового действия для программного приложения разрешает новое голосовое действие для программного приложения.

15. Компьютерно-реализуемый способ по п. 1, содержащий этапы, на которых:

5 принимают посредством системы голосовых действий от разработчика, связанного с программным приложением, запрос на отмену развертывания нового голосового действия для программного приложения; и

в ответ на этот запрос отменяют развертывание нового голосового действия для программного приложения в ответ на данный запрос, при этом отмена развертывания

10 нового голосового действия для программного приложения запрещает новое голосовое действие для программного приложения.

16. Компьютерно-реализуемый способ по п. 1, содержащий этапы, на которых:

принимают посредством системы голосовых действий от разработчика, связанного с программным приложением, запрос на разрешение тестирования нового голосового 15 действия, каковой запрос указывает одно или более устройств, для которых должно быть разрешено новое голосовое действие; и

в ответ на этот запрос разрешают новое голосовое действие для одного или более указанных устройств, при этом новое голосовое действие запрещено для устройств, которые не включены в указанные устройства.

20 17. Компьютерно-реализуемый способ по п. 1, содержащий этапы, на которых:

принимают посредством системы голосовых действий указание относительно 25 высказывания пользователя, полученного конкретным устройством, имеющим установленное программное приложение;

определяют посредством системы голосовых действий, что транскрипция

высказывания пользователя соответствует одному или более триггерным элементам, связанным с контекстным намерением голосового действия; и

в ответ на это определение предоставляют посредством системы голосовых действий конкретному устройству контекстное намерение голосового действия, тем самым запрашивая, чтобы программное приложение, установленное на конкретном устройстве,

30 выполнило одну или более операций нового голосового действия.

18. Система голосовых действий, выполненная с возможностью назначения новых голосовых действий для программного приложения, которое ранее было установлено на пользовательских устройствах, отличающихся от системы голосовых действий, причем голосовое действие задает одну или более операций, выполняемых в

35 программном приложении, при этом система голосовых действий содержит один или более компьютеров и одно или более запоминающих устройств, хранящих инструкции, которые вызывают при их исполнении одним или более компьютерами выполнение одним или более компьютерами операций, содержащих:

прием данных, указывающих новое голосовое действие для программного

40 приложения, каковые данные содержат (1) одну или более операций для выполнения нового голосового действия, (2) один или более триггерных элементов для активации нового голосового действия и (3) контекст, указывающий статус пользовательского устройства или программного приложения, когда новое голосовое действие разрешено;

генерирование пассивной структуры данных голосового действия для программного

45 приложения на основе, по меньшей мере, принятых данных, при этом пассивная структура данных голосового действия содержит данные, которые при приеме программным приложением запрашивают, чтобы программное приложение выполнило одну или более операций нового голосового действия;

связывание пассивной структуры данных голосового действия с контекстом и с одним или более триггерными элементами для нового голосового действия;

при этом система голосовых действий выполнена с возможностью:

принимать от пользовательского устройства указание относительно (1) высказывания

5 пользователя, полученного устройством, на котором установлено программное приложение, и (2) контекстной информации, указывающей статус пользовательского устройства, имеющего установленное программное приложение, или программного приложения, установленного на пользовательском устройстве;

10 определять, что (1) транскрипция высказывания пользователя соответствует одному или более триггерным элементам, связанным с пассивной структурой данных голосового действия, и что (2) контекстная информация удовлетворяет контексту для нового голосового действия; и

15 в ответ на это определение подавать в пользовательское устройство пассивную структуру данных голосового действия, тем самым запрашивая, чтобы программное приложение, установленное на устройстве, выполнило одну или более операций нового голосового действия.

19. Долговременное машиночитаемое запоминающее устройство, хранящее инструкции, исполняемые одним или более компьютерами, которые при таком исполнении вызывают выполнение одним или более компьютерами операций,

20 содержащих:

прием, посредством системы голосовых действий, данных, указывающих новое голосовое действие для программного приложения, отличающегося от упомянутой системы голосовых действий, каковые данные содержат (1) одну или более операций для выполнения нового голосового действия, (2) один или более триггерных элементов 25 для активации нового голосового действия и (3) контекст, указывающий статус пользовательского устройства или программного приложения, когда новое голосовое действие разрешено;

30 генерирование, посредством системы голосовых действий, пассивной структуры данных голосового действия для программного приложения на основе, по меньшей мере, принятых данных, при этом пассивная структура данных голосового действия содержит данные, которые при приеме программным приложением запрашиваются, чтобы программное приложение выполнило одну или более операций нового голосового действия;

35 связывание, посредством системы голосовых действий, пассивной структуры данных голосового действия с контекстом и с одним или более триггерными элементами для нового голосового действия;

при этом система голосовых действий выполнена с возможностью:

40 принимать от пользовательского устройства указание относительно (1) высказывания пользователя, полученного устройством, на котором установлено программное приложение, и (2) контекстной информации, указывающей статус пользовательского устройства, имеющего установленное программное приложение, или программного приложения, установленного на пользовательском устройстве;

45 определять, что (1) транскрипция высказывания пользователя соответствует одному или более триггерным элементам, связанным с пассивной структурой данных голосового действия, и что (2) контекстная информация удовлетворяет контексту для нового голосового действия; и

в ответ на это определение подавать в пользовательское устройство пассивную структуру данных голосового действия, тем самым запрашивая, чтобы программное

приложение, установленное на устройстве, выполнило одну или более операций нового голосового действия.

5

10

15

20

25

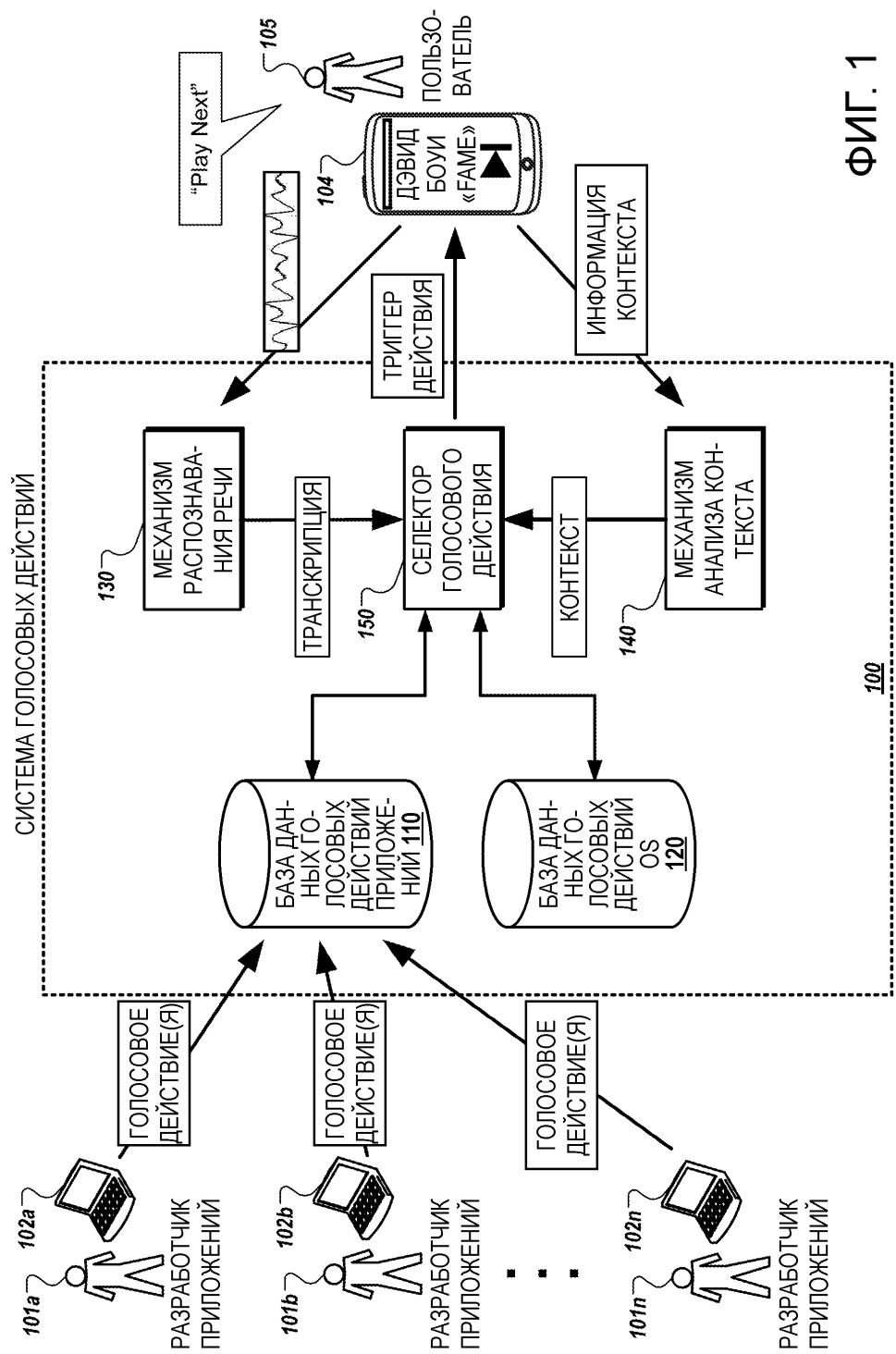
30

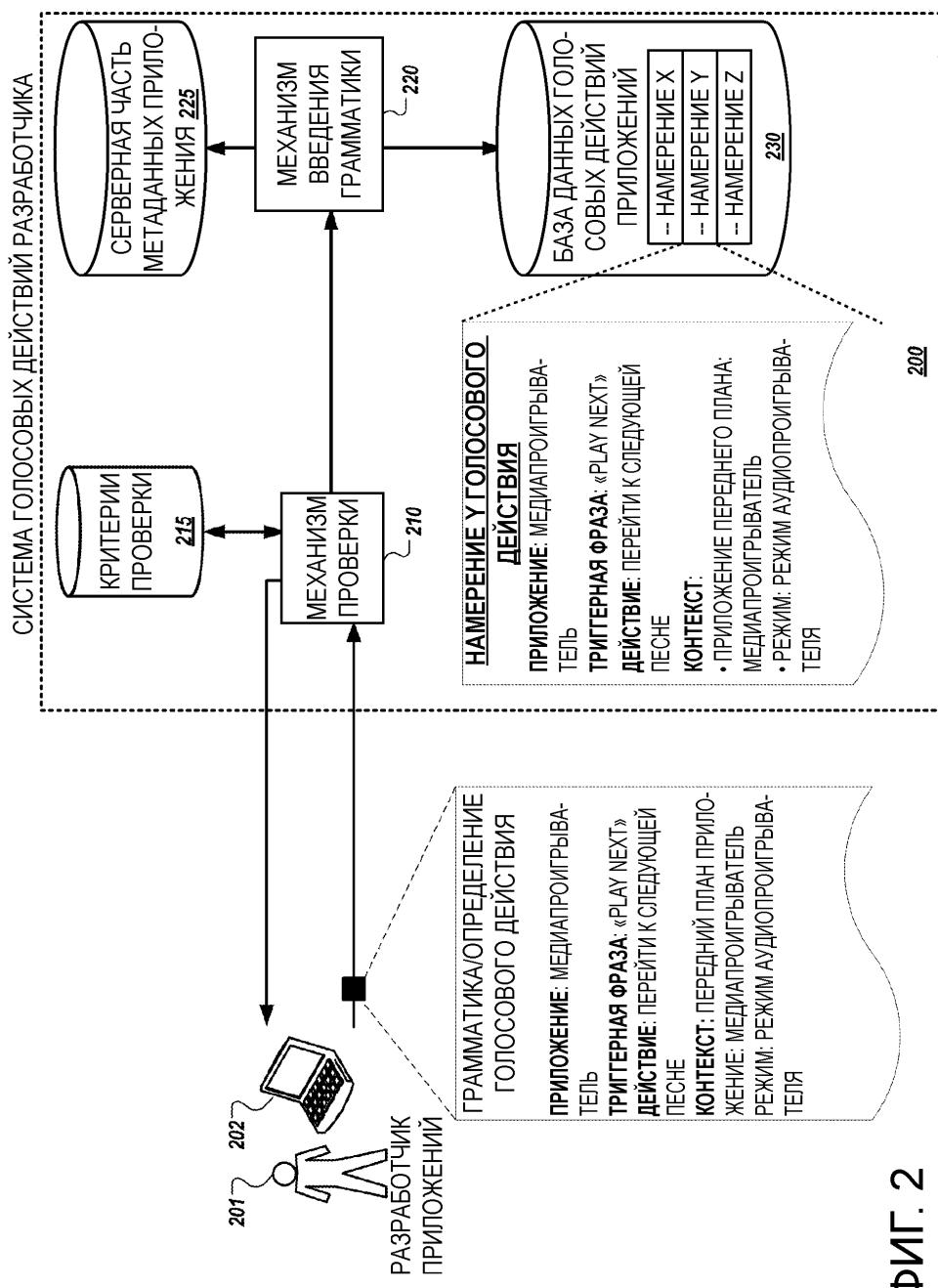
35

40

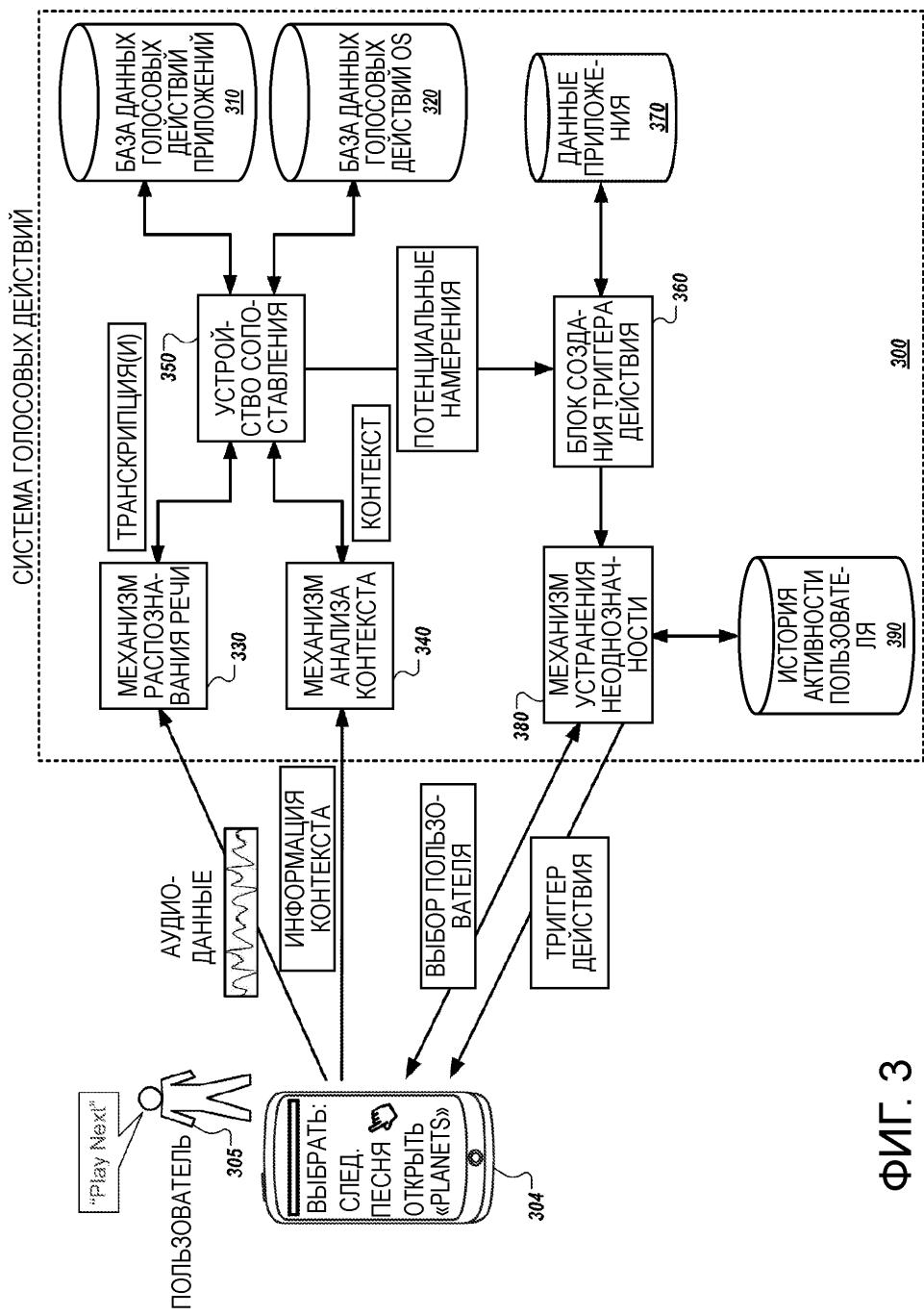
45

ФИГ. 1



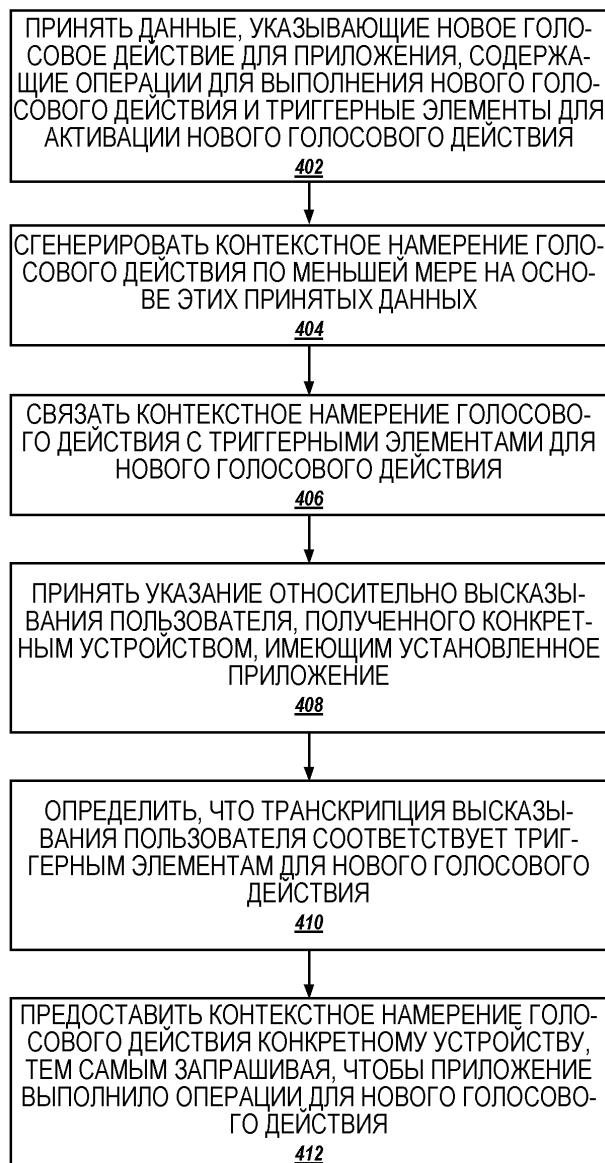


ФИГ. 2



ΦΙΓ. 3

400



ФИГ. 4