



morfeusys 21 сентября 2014 в 22:31

Сверхбыстрое распознавание речи без серверов на реальном приме

Разработка под Android

Tutorial

Как создать умный дом с голосовым управлением



В этой статье я подробно расскажу и покажу, как правильно и быстро прикрутить распознавание русской речи на движке Pocketsphinx (для порт OpenEars) на реальном ~~Hello World~~ примере управления домашней техникой.

Почему именно домашней техникой? Да потому что благодаря такому примеру можно оценить ту **скорость и точность**, которой можно дог при использовании **полностью локального** распознавания речи без серверов типа *Google ASR* или *Яндекс SpeechKit*.

К статье я также прилагаю все исходники программы и саму сборку под Android.

С чего вдруг?

Наткнувшись недавно на статью о прикручивании Яндекс SpeechKit-а к iOS приложению, я задал вопрос автору, почему для своей програм захотел использовать именно серверное распознавание речи (по моему мнению, это было излишним и приводило к некоторым проблемам что получил встречный вопрос о том, не мог бы я поподробней описать применение альтернативных способов для проектов, где нет необходимости распознавать что угодно, а словарь состоит из конечного набора слов. Да еще и с примером практического применения...

Зачем нам что-то еще кроме Яндекса и Google?

В качестве того самого «практического применения» я выбрал тему **голосового управления умным домом**.

Почему именно такой пример? Потому что на нем можно увидеть те несколько преимуществ полностью локального распознавания речи по сравнению с использованием облачных решений. А именно:

- **Скорость** — мы не зависим от серверов и поэтому не зависим от их доступности, пропускной способности и т.п. факторов
- **Точность** — наш движок работает только с тем словарем, который интересует наше приложение, повышая тем самым качество распознавания
- **Стоимость** — нам не придется платить за каждый запрос к серверу
- **Голосовая активация** — как дополнительный бонус к первым пунктам — мы можем постоянно «слушать эфир», не тратя при этом свс трафик и не нагружая сервера

[Примечание](#)

Так Android же умеет распознавать речь без интернета!

Да-да... Только на JellyBean. И только с полуметра, не более. И это распознавание — это та же диктовка, только с использованием гораздо меньшей модели. Так что управлять ею и настраивать ее мы тоже не можем. И что она вернет нам в следующий раз — неизвестно. Хотя ди СМС-ок в самый раз!

Что будем делать?



Будем реализовывать голосовой пульт управления домашней техникой, который будет работать точно и быстро, с нескольких метров и даже в дешевом термозонном хламе очень недорогих Android смартфонах, планшетах и часах.

Логика будет простой, но очень практичной. Активируем микрофон и произносим одно или несколько названий устройств. Приложение их распознает и включает-выключает их в зависимости от текущего состояния. Либо получает от них состояние и произносит его приятным же голосом. Например, текущая температура в комнате.

Микрофон будем активировать или голосом, или нажатием на иконку микрофона, или даже просто положив руку на экран. Экран в свою очередь может быть и полностью выключенным.

Вариантов практического применения масса

На видео показано, что получилось в итоге. Далее же речь пойдет о технической реализации с выдержками из реально работающего кода немного теории.

Что такое Pocketsphinx

CMU Sphinx

Pocketsphinx — это движок распознавания с открытым исходным кодом под Android. У него также имеется порт под iOS, WindowsPhone, и даже JavaScript.

Он позволит нам запустить распознавание речи прямо на устройстве и при этом настроить его именно под наши задачи. Также он предлагает функцию голосовой активации «из коробки» (см далее).

Мы сможем «скормить» движку распознавания русскую языковую модель (вы можете найти ее в исходниках) и грамматику пользовательски запросов. Это именно то, что будет распознавать наше приложение. Ничего другого оно распознать не сможет. А следовательно, практически никогда не выдаст что-то, чего мы не ожидаем.

Грамматика JSGF

Pocketsphinx также может работать по статистической модели языка, что позволяет распознавать спонтанную речь, не описываемую контекстной свободной грамматикой. Но для нашей задачи это как раз не нужно. Наша грамматика будет состоять только из названий устройств. После процесса распознавания Pocketsphinx вернет нам обычную строку текста, где устройства будут идти один за другим.

```
#JSGF V1.0;
grammar commands;
public <command> = <commands>+;
<commands> = лапма | монитор | температура;
```

Знак *плюса* обозначает, что пользователь может назвать не одно, а несколько устройств подряд.

Приложение получает список устройств от контроллера умного дома (см далее) и формирует такую грамматику в классе Grammar.

Транскрипции

i:	I	U	u:	Iə	eɪ	ɪ	↗
eat	big	cook	school	beer	day		
e	ə	ɜ:	ɔ:	ʊə	ɔɪ	əʊ	
ten	the	work	small	sure	boy	old	
æ	ʌ	ɑ:	ɒ	eə	aɪ	aʊ	
map	cup	father	dog	hair	wine	house	
p	b	t	d	tʃ	dʒ	k	g
pen	bag	ticket	door	chair	job	key	glass
f	v	θ	ð	s	z	ʃ	ʒ
from	video	three	this	sister	zoo	shop	television
m	n	ŋ	h	l	r	w	j
morning	name	sing	home	live	read	window	yes

Грамматика описывает то, **что может говорить пользователь**. Для того, чтобы Pocketsphinx знал, **как** он это будет произносить, необходимо каждого слова из грамматики написать, как оно звучит в соответствующей языковой модели. То есть *транскрипцию* каждого слова. Это называется *словарь*.

Транскрипции описываются с помощью специального синтаксиса. Например:

```
умный  uu m n ay j
дом     d oo m
```

В принципе, ничего сложного. Двойная гласная в транскрипции обозначает ударение. Двойная согласная — мягкую согласную, за которой гласная. Все возможные комбинации для всех звуков русского языка можно найти в самой языковой модели.

Понятно, что заранее описать все транскрипции в нашем приложении мы не можем, потому что мы не знаем заранее тех названий, которые пользователь даст своим устройствам. Поэтому мы будем генерировать «на лету» такие транскрипции по некоторым правилам русской фонетики. Для этого можно реализовать вот такой класс `PhonMapper`, который сможет получать на вход строчку и генерировать для нее правильную транскрипцию.

Голосовая активация

Это возможность движка распознавания речи все время «слушать эфир» с целью реакции на заранее заданную фразу (или фразы). При этом другие звуки и речь будут отбрасываться. Это не то же самое, что описать грамматику и просто включить микрофон. Приводить здесь теорию этой задачи и механику того, как это работает, я не буду. Скажу лишь только, что недавно программисты, работающие над Pocketsphinx, реализовали такую функцию, и теперь она доступна «из коробки» в API.

Одно стоит упомянуть обязательно. Для активационной фразы нужно не только указать транскрипцию, но и подобрать подходящее **значение порога чувствительности**. Слишком маленькое значение приведет к множеству ложных срабатываний (это когда вы не говорили активационную фразу, а система ее распознает). А слишком высокое — к невосприимчивости. Поэтому данная настройка имеет особую важность. Пример диапазон значений — от $1e-1$ до $1e-40$ в зависимости от активационной фразы.

Активация по датчику приближения

Запускаем распознавание

Pocketsphinx предоставляет удобный API для конфигурирования и запуска процесса распознавания. Это классы *SpeechRecognizer* и *SpeechRecognizerSetup*.

Вот как выглядит конфигурация и запуск распознавания:

```
PhonMapper phonMapper = new PhonMapper(getAssets().open("dict/ru/hotwords"));
Grammar grammar = new Grammar(names, phonMapper);
grammar.addWords(hotword);
DataFiles dataFiles = new DataFiles(getPackageName(), "ru");
File hmmDir = new File(dataFiles.getHmm());
File dict = new File(dataFiles.getDict());
File js gf = new File(dataFiles.getJsgf());
copyAssets(hmmDir);
saveFile(jsgf, grammar.getJsgf());
saveFile(dict, grammar.getDict());
mRecognizer = SpeechRecognizerSetup.defaultSetup()
    .setAcousticModel(hmmDir)
    .setDictionary(dict)
    .setBoolean("--remove_noise", false)
    .setKeywordThreshold(1e-7f)
    .getRecognizer();
mRecognizer.addKeyphraseSearch(KWS_SEARCH, hotword);
mRecognizer.addGrammarSearch(COMMAND_SEARCH, jsgf);
```

Здесь мы сперва копируем все необходимые файлы на диск (Pocketsphinx требует наличия на диске акустической модели, грамматики и с транскрипциями). Затем конфигурируется сам движок распознавания. Указываются пути к файлам модели и словаря, а также некоторые параметры (порог чувствительности для активационной фразы). Далее конфигурируется путь к файлу с грамматикой, а также активационн фраза.

Как видно из этого кода, один движок конфигурируется сразу и для грамматики, и для распознавания активационной фразы. Зачем так дел: Для того, чтобы мы могли быстро переключаться между тем, что в данный момент нужно распознавать. Вот как выглядит запуск процесса распознавания активационной фразы:

```
mRecognizer.startListening(KWS_SEARCH);
```

А вот так — распознавание речи по заданной грамматике:

```
mRecognizer.startListening(COMMAND_SEARCH, 3000);
```

Второй аргумент (необязательный) — количество миллисекунд, после которого распознавание будет автоматически завершаться, если ник ничего не говорит.

Как видите, можно использовать только один движок для решения обеих задач.

Как получить результат распознавания

Чтобы получить результат распознавания, нужно также указать слушателя событий, имплементирующего интерфейс *RecognitionListener*. У него есть несколько методов, которые вызываются rocketsphinx-ом при наступлении одного из событий:

- **onBeginningOfSpeech** — движок услышал какой-то звук, может быть это речь (а может быть и нет)
- **onEndOfSpeech** — звук закончился
- **onPartialResult** — есть промежуточные результаты распознавания. Для активационной фразы это значит, что она сработала. Аргумент *Hypothesis* содержит данные о распознавании (строка и score)
- **onResult** — конечный результат распознавания. Этот метод будет вызван после вызова метода *stop* у *SpeechRecognizer*. Аргумент *Hypothesis* содержит данные о распознавании (строка и score)

Реализуя тем или иным способом методы `onPartialResult` и `onResult`, можно изменять логику распознавания и получать окончательный результат. Вот как это сделано в случае с нашим приложением:

```
@Override
public void onEndOfSpeech() {
    Log.d(TAG, "onEndOfSpeech");
    if (mRecognizer.getSearchName().equals(COMMAND_SEARCH)) {
        mRecognizer.stop();
    }
}

@Override
public void onPartialResult(Hypothesis hypothesis) {
    if (hypothesis == null) return;
    String text = hypothesis.getHypstr();
    if (KWS_SEARCH.equals(mRecognizer.getSearchName())) {
        startRecognition();
    } else {
        Log.d(TAG, text);
    }
}

@Override
public void onResult(Hypothesis hypothesis) {
    mMicView.setBackgroundResource(R.drawable.background_big_mic);
    mHandler.removeCallbacks(mStopRecognitionCallback);
    String text = hypothesis != null ? hypothesis.getHypstr() : null;
    Log.d(TAG, "onResult " + text);

    if (COMMAND_SEARCH.equals(mRecognizer.getSearchName())) {
        if (text != null) {
            Toast.makeText(this, text, Toast.LENGTH_SHORT).show();
            process(text);
        }
        mRecognizer.startListening(KWS_SEARCH);
    }
}
```

Когда мы получаем событие `onEndOfSpeech`, и если при этом мы распознаем команду для выполнения, то необходимо остановить распознавание, после чего сразу будет вызван `onResult`.

В `onResult` нужно проверить, что только что было распознано. Если это команда, то нужно запустить ее на выполнение и переключить движок распознавания активационной фразы.

В `onPartialResult` нас интересует только распознавание активационной фразы. Если мы его обнаруживаем, то сразу запускаем процесс распознавания команды. Вот как он выглядит:

```
private synchronized void startRecognition() {
    if (mRecognizer == null || COMMAND_SEARCH.equals(mRecognizer.getSearchName())) return;
    mRecognizer.cancel();
    new ToneGenerator(AudioManager.STREAM_MUSIC, ToneGenerator.MAX_VOLUME).startTone(ToneGenerator.TONE_CDMA_PIP, 200);
    post(400, new Runnable() {
        @Override
        public void run() {
            mMicView.setBackgroundResource(R.drawable.background_big_mic_green);
            mRecognizer.startListening(COMMAND_SEARCH, 3000);
            Log.d(TAG, "Listen commands");
            post(4000, mStopRecognitionCallback);
        }
    });
}
```

Здесь мы сперва играем небольшой сигнал для оповещения пользователя, что мы его услышали и готовы к его команде. На это время микрофон должен быть выключен. Поэтому мы запускаем распознавание после небольшого таймаута (чуть больше, чем длительность сигнала, чтобы услышать его эхо). Также запускается поток, который остановит распознавание принудительно, если пользователь говорит слишком долго. В данном случае это 3 секунды.

Как превратить распознанную строку в команды

Ну тут все уже специфично для конкретного приложения. В случае с нашим примером, мы просто вытягиваем из строки названия устройств

ищем по ним нужное устройство и либо меняем его состояние с помощью HTTP запроса на контроллер умного дома, либо сообщаем его текущее состояние (как в случае с термостатом). Эту логику можно увидеть в классе `Controller`.

Как синтезировать речь

Синтез речи — это операция, обратная распознаванию. Здесь наоборот — нужно превратить строку текста в речь, чтобы ее услышал пользователь.

В случае с термостатом мы должны заставить наше Android устройство произнести текущую температуру. С помощью API `TextToSpeech` это сделать довольно просто (спасибо гуглу за прекрасный женский TTS для русского языка):

```
private void speak(String text) {
    synchronized (mSpeechQueue) {
        mRecognizer.stop();
        mSpeechQueue.add(text);
        HashMap<String, String> params = new HashMap<String, String>(2);
        params.put(TextToSpeech.Engine.KEY_PARAM_UTTERANCE_ID, UUID.randomUUID().toString());
        params.put(TextToSpeech.Engine.KEY_PARAM_STREAM, String.valueOf(AudioManager.STREAM_MUSIC));
        params.put(TextToSpeech.Engine.KEY_FEATURE_NETWORK_SYNTHESIS, "true");
        mTextToSpeech.speak(text, TextToSpeech.QUEUE_ADD, params);
    }
}
```

Скажу наверное банальность, но **перед процессом синтеза нужно обязательно отключить распознавание**. На некоторых устройствах (например, все самсунги) вообще невозможно одновременно и слушать микрофон, и что-то синтезировать.

Окончание синтеза речи (то есть окончание процесса говорения текста синтезатором) можно отследить в слушателе:

```
private final TextToSpeech.OnUtteranceCompletedListener mUtteranceCompletedListener = new TextToSpeech.OnUtteranceCompletedListener() {
    @Override
    public void onUtteranceCompleted(String utteranceId) {
        synchronized (mSpeechQueue) {
            mSpeechQueue.poll();
            if (mSpeechQueue.isEmpty()) {
                mRecognizer.startListening(KWS_SEARCH);
            }
        }
    }
};
```

В нем мы просто проверяем, нет ли еще чего-то в очереди на синтез, и включаем распознавание активационной фразы, если ничего больше нет.

И это все?

Да! Как видите, быстро и качественно распознать речь прямо на устройстве совсем несложно, благодаря наличию таких замечательных как `Rocketsphinx`. Он предоставляет очень удобный API, который можно использовать в решении задач, связанных с распознаванием голосовых команд.

В данном примере мы прикрутили распознавание к вполне кокретной задаче — *голосовому управлению устройствами умного дома*. За счет локального распознавания мы добились очень высокой скорости работы и минимизировали ошибки.

Понятно, что тот же код можно использовать и для других задач, связанных с голосом. Это не обязательно должен быть именно умный дом.

Все исходники, а также саму сборку приложения вы можете найти в репозитории на [GitHub](#).

Также на моем канале в [YouTube](#) вы можете увидеть некоторые другие реализации голосового управления, и не только системами умных домов.

Теги: android, умный дом, распознавание речи, ios development, android development, голосовое управление, голосовой движок

↑ +179 ↓ 1617 271k 84



63,0

Карма

0,0

Рейтинг

53

Подписчики

Дмитрий Че @morfeusys

Программист

Сайт

Поделиться публикацией

ПОХОЖИЕ ПУБЛИКАЦИИ

- 10 декабря 2018 в 13:16

Система управления умным домом на коленке: Tarantool

+34

12,2k

140

32
- 9 января 2018 в 22:22

Безлимитное распознавание речи. Или как я перевожу в боте голосовые сообщения в текст

+37

20,3k

256

7
- 15 июля 2014 в 01:17






Никогда не «не делай» того, о чем пожалеешь или умный дом с CCU.IO

+78


110k

759

86

ВАКАНСИИ		Мой к
	Android разработчик Asapp · Москва · Возможна удаленная работа	от 150000 до 2000
	Android developer (Kotlin) Frogogo · Москва · Возможна удаленная работа	от 800
	Android developer Blogman · Оренбург	от 20000 до 1100
	Старший Android разработчик Dialog · Москва	от 160000 до 3000
	Android разработчик KUPIBILET.RU · Санкт-Петербург	от 1000
Все вакансии		


Комментарии 84

-  aivs

21 сентября 2014 в 23:07

#


↑

Спасибо, очень круто, как раз планировал под iOS своему приложению управления умным домом добавить поддержку голоса!
-  BeIBES

21 сентября 2014 в 23:13

#

↑







































































А для чего используется этот контроллер VeraLite?
-  morfeusys

21 сентября 2014 в 23:24

#

↑

Так для управления оконечными устройствами. Он формирует сеть ZWave, к которой подключаются устройства. А далее контроллер получает HT запросы от приложения и транслирует их в команды ZWave протокола.

-  **lightcaster** 21 сентября 2014 в 23:35   
- «Телефончик, слушай, отправь смсочку Лене и скажи задержись на пять минут.»
- Голосовое управление станет обыденностью если дивайс сможет распознать подобные вещи, сказанные бегло и без напряжения.
-  **morfeusys** 21 сентября 2014 в 23:48     
- Статья «немножко» о другом))
-  **lightcaster** 22 сентября 2014 в 07:09     
- О, заминусили :)
- Я имел ввиду что вы говорите довольно «роботизированным» голосом, с четкими паузами и разделениями слов. Более того, короткими командами. Это не очень практично в обычной жизни, где человек говорит бегло.
-  **morfeusys** 22 сентября 2014 в 08:59     
- Практично? Роботизированным? Дома удобно быстро сказать два-три слова типа «телевизор, кондей, свет на кухне». А не пространными объяснениями типа «уважаемый умный дом, не могли бы вы включить мне...» И далее.
- И чтобы отработало моментально. Я говорю нормальным голосом без особых пауз. Нам нужна четкая голосовая машина, а не «собеседник-чатбот». Вот я о таком подходе и рассказал.
-  **rPman** 22 сентября 2014 в 09:20     
- Вы не поверите (с)
- Голосовой ассистент Дуся: 4pda.ru/forum/index.php?showtopic=561485 там же есть несколько видеодемонстраций
- К примеру фразы 'Дуся, напиши смску леночке я задержусь на пять минут' или 'Дуся, разбуди меня через три часа' будут распознаны и обработаны. Работает offline! так как голосовой движок от google уже работает без интернета, правда чуть медленней чем онлайн вариант.
-  **morfeusys** 22 сентября 2014 в 10:11     
- www.youtube.com/watch?v=B3ygkiXEurY
www.youtube.com/watch?v=76YB4JGOblo
-  **mickvav** 22 сентября 2014 в 14:31     
- Почитал эту тему на 4pda —
- Как запустить Дусю из сторонней программы с определенной фразой
- Для этого нужно запустить Broadcast с Action com.dusiassistant.INPUT и в Extras передать строку с именем text
- и подумал, что это же отличный вектор для атаки — фигачим в какую-нибудь говноигрушку «отправь смс на <короткий номер> с текстом ...» и и дусеводов...
-  **morfeusys** 22 сентября 2014 в 15:14     
- Угу, при том что Дуся обязательно при этом переспросит голосом владельца, прежде чем что-то отправлять))
-  **rPman** 22 сентября 2014 в 23:18     
- Ну уязвимости то есть :(
- Пока я игрался с Дусей, обнаружил что при любых включенных механизмах активации на экране блокировки (хоть голосом, хоть потряхиванием, хоть проведением рукой) есть возможность запустить любое приложение (это отключается, как и варианты с навигацией курсами валют и т.п., реализованные через браузер) сказав 'Открой xxx' например браузер, он будет запущен и позволит с ним работать снятия блокировки.
-  **morfeusys** 23 сентября 2014 в 00:08     
- Уязвимости? Ээ... Так поставьте графический ключик. Хотя Если ваш телефон в руки взял злоумышленник — то пенять тут на ассистента толку, ваши данные и без него станут доступны.
- Кстати все виды активации можно вообще выключить. Так что об «уязвимостях» тут говорить неуместно.
-  **rPman** 23 сентября 2014 в 00:33     
- Я тестировал именно при блокировке графическим ключом.
- И да, это уязвимость, идеологическая, потому как при закрытии большинства механизмов активации, остается единственный надежный но неудобный — по запуску приложения.
-  **morfeusys** 23 сентября 2014 в 09:03     

Есть еще по тапу на виджет. Или с бт гарнитуры. Или из шторки — быстрый старт. У Дуси активаций масса. Читайте внимательно! Если у вас разблокируется при графическом ключе — то это уже проблемы вашей прошивки.



NightGhost 22 сентября 2014 в 18:03



То, о чем вы пишете, это уже не просто распознавание речи, а интеллектуальный анализ распознанного. Разные вещи.



lightcaster 22 сентября 2014 в 18:18



Вы не поверите, насколько сильно одно связано с другим.



NightGhost 22 сентября 2014 в 18:23



Связано то связано, но вещи разные



morfeusys 22 сентября 2014 в 18:43



Это вещи на стыке разных технологий. Задача процессинга текста по сути никак не связана с распознаванием. Также как и распознавание с процессингом. Но при реализации в единой парадигме можно решать обе задачи более эффективно.



Zelgadis 30 января 2015 в 01:16



Мимо проходил: wit.ai/



L3n1n 21 сентября 2014 в 23:59



Спасибо за статью!

Но вера тут никак не обязательна. После распознавания команды можно отправлять запрос на любой контроллер (в моем случае Fibaro).



morfeusys 22 сентября 2014 в 00:05



Да конечно не обязательна. Статья о том, как можно прикрутить такое голосовое управление к чему угодно. У меня есть вера3, к нему и прикрути. Просто у фибары с верой могут быть разные http протоколы. Поэтому не могу сказать, будет ли работать с фибарой тот же код.



samodum 22 сентября 2014 в 00:28



Отличная статья! Спасибо!



byria 22 сентября 2014 в 00:40



Уважаемый автор, подскажите, по нескольким вопросам:

1. Можно ли использовать данный API не с мобильным OS, к примеру с Linux установленным на мини PC.
2. Какие телефоны из дешевых можно использовать на Android (было что-то сказано против Samsung)?
3. Можно ли обойтись Vega или аналогом в 1 единственном числе, если у вас дом в несколько этажей и 10 комнат, и к примеру у вас в каждой комнате дешевому телефону (как приемник) и как минимум в комнате до 10 устройств (розеток) под управление.
4. Из 3 вопроса выходит 4 — как локально определять однотипные названия, к примеру жалюзи? Запоминать что в этой комнате есть жалюзи и нужно называть ЖАЛЮЗИ №3 — не удобно. Интересует самоопределение по нахождению с привязкой через конкретный телефон в конкретной комнате.
5. К примеру я сделал несколько устройств с использованием микроконтроллеров (Arduino с wifi) сам и еще много всего, узкое место — роутер с его пропускной способностью? Сложно ли подружить DIY устройства?
6. Есть ли оболочка, которая устанавливается на 1 сервер, к которому можно подключиться в графическом режиме через планшет к примеру, указат куплена новая умная розетка, задать ей имя, указать привязку к конкретному телефону (к примеру телефон № 4 — зал) и по принципу Plug&Play вкл за 5 минут?



Griffonn 22 сентября 2014 в 01:03



4. Не ориентируюсь в конкретных контроллерах, но разве нельзя передать вместе с фразой идентификатор телефона дополнительным параметром пусть контроллер реагирует соответственно?



morfeusys 22 сентября 2014 в 01:09





Ух как вопросов много. Сразу видно, что вопросом владеете))





1. Можно. На сайте [rocketsphinx](http://rocketsphinx.com) вы можете увидеть один из подобных проектов. Кстати на том же raspberry или cubieboard нормально ставится андроид. Придется поэкспериментировать с микрофонами и звуковыми платами.
2. Да в принципе любые. Ну например мегафон клик, старые htc hero и тп. Насчет самсунга — ничего против. Просто указал, что на них нельзя одновременно и слушать и синтезировать звук. Кстати у самсунгов одни из лучших микрофонов. Даже на самых дешевых моделях.
3. Ну начнем с того, что умный дом — это далеко не только розетки)) Насчет самой сети zwave — в принципе можно обойтись при особом расчете. Дело в том, что zwave — это самоорганизующаяся сеть, с функцией «черезсеть», т.е. Каждое устройство является ретранслятором радио сигнала. Если правильно расположить устройства так, чтобы каждая пара была в области видимости на несколько метров, то возможно обойтись и одним контроллером. Не будем также забывать о возможности использовать гораздо более дешевые usb контроллеры, подключаемые к ПК.
4. Это уже зависит от логики работы приложения на смартфоне (пульт). Исходники я приложил — можно экспериментировать)) В одном из проектов вообще делали автоматическое определение местоположения андроида по bluetooth маячкам по всему дому — работало хорошо.
5. Несложно, если вы используете те же протоколы. Но для разношерстных устройств в одной сети интернета вещей я предлагаю использовать т

orenhab — он может объединить все устройства на более 100 протоколах. Почитайте в моей предыдущей статье.


6. Это реализовано в тех же контроллерах Vera. Посмотрите мое видео до конца.

 **Neuronix** 7 октября 2014 в 11:37    

Вы сейчас приблизительно описываете то, что я делаю :)

 **morfeusys** 7 октября 2014 в 12:00    


Что конкретно? интересно — давайте пообщаемся

 **Neuronix** 7 октября 2014 в 12:32    

Да, собственно, всё, кроме пункта 5. В качестве контроллера используется обычный ПК с набором донглов.

Если интересно, то код можно найти у меня в профиле в GitHub. Проект разбит на 3 приложения — ядро системы, веб-интерфейс и термин Android.

Второе и третье в зачаточной стадии, как раз пытаюсь прикрутить сфинкс к андроиду.

 **byria** 22 сентября 2014 в 01:22  

Спасибо за ответы! Но вопросом не владею)) как заметили, только только к роботу прикручивать буду голосовые команды и движок выбираю, ранее использовался другой, самопальный. Параллельно и дома хочется что-то сделать умного.




4. Сама по себе статья про определение местоположения, если возможно, была бы не менее интересна на основе вашего опыта.

 **morfeusys** 22 сентября 2014 в 01:27    

Хм... А это можно... Если будет интересно. Это тема про bluetooth low energy (4-я версия).

 **mbait**  22 сентября 2014 в 02:00  

Я один из соавторов pocketsphinx-android, могу ответить на некоторые вопросы. Есть одна проблема, с которой у всех возникают вопросы, и которая пропущена в этой статье — ресурсы приложения. Суть проблемы в том, что pocketsphinx требует путь, чтобы прочитать файл, а для Android assets возможно получить только InputStream. Поэтому для демонстрационного приложения я написал ant-скрипт, который вычисляет .md5 суммы всех файлов из директории assets/sync и создаёт список, который потом используется для синхронизации файлов между .apk и внешней памятью (обычно, SD-карта). Поэтому, если вы собираете проект через ADT плагин, то убедитесь, что включены соответствующие Ant-шаги в процесс сборки. Впрочем, есть подробное описание, несмотря на существование которого, стабильно раз в неделю кто-нибудь да и напишет в форуме, что у него проблемы с ресурсами.

 **byria** 22 сентября 2014 в 03:02    

У меня есть вопрос иного характера.



Не получится ли так, что подсев на данный движок, со временем он станет платным? Есть такие планы?

 **mbait** 22 сентября 2014 в 03:04    

Нет, код полностью открытый. Библиотеку можно использовать в коммерческих и закрытых проектах.

 **Neuronix** 7 октября 2014 в 11:44    

Т.е. то что тут лежит в assets не подтягивается pocketsphinx?

 **mbait** 7 октября 2014 в 12:10    




Нет, нужно положить в assets/sync и сгенерировать .md5 для каждого файла и контрольный файл со списком ресурсов. В pocketsphinx-android это делается автоматически через ant-скрипт, но вы можете использовать любой способ. Если вы хотите, чтобы все файлы из assets/ синхронизировались, то нужно изменить название директории в pocketsphinx-android.

 **ClusterM** 8 октября 2014 в 01:15    

Не совсем понял. В какой момент времени происходит синхронизация с внешней памятью? И где именно они в итоге лежат? Ни во внутренней памяти ни на SD карте не нахожу, хотя демонстрационное приложение работает отлично.

 **ClusterM** 8 октября 2014 в 01:28    

Прошу прощения, разобрался в коде примера, вопрос отпал.

 **v0id** 17 февраля 2016 в 14:05    

Здравствуйте. Я только начинаю разбираться с программированием под android.

Пример из статьи не заработал, взял пример с сайта библиотеки — заработал, но там только английский язык

Можете помочь, как добавить в пример с сайта русский для активационной фразы и для распознавания двух команд?

 **Afganec** 17 февраля 2016 в 17:47    

Могу помочь, пишите в личку



lless 22 сентября 2014 в 09:50



А можно ли прикрутить управление телевизором по голосу? телевизор включить, москва24, Громкость плюс 10? По ИК или WiFi(SmartTV)



morfeusys 22 сентября 2014 в 10:11



www.youtube.com/watch?v=76YB4JGOblo



axeax 22 сентября 2014 в 10:49



Холодно у вас дома)



Vinchi 22 сентября 2014 в 11:36



может есть статистика по зависимости ложных срабатываний от количества слов и грамматик?



morfeusys 22 сентября 2014 в 11:40



Грамматики тут не принимают участия. Если вы про активацию голосом. В активации голосом нужно подбирать подходящее значение порога. Плюс еще важно, что активационная фраза не должна быть короткой. Примерно 2 слова по 2-3 слога в каждом. Необходимо также очень аккуратно писать транскрипции.



utya 22 сентября 2014 в 11:54



Круто, сейчас тоже прикручиваю. Но с помощью 1 Sheeld и не совсем удачно работает распознавание голоса. Какова дальность определение? Дума пристроить к андроид девайсу микрофон ШОРОХ



pelment 22 сентября 2014 в 15:40



У меня был очень негативный опыт работы с rockeysphinx под Ubuntu и ROS. Нужно было распознавание, которое реально совместить с ROS, быст работающее и не как можно меньше зависящее от внешних условий, дикторонезависимое. Мне так и не удалось настроить rocketsphinx на хорошее распознавание, постоянные ошибки были. После такой оптимистичной статьи попробую еще раз при случае — может, у меня руки не из того места оказались.



mbait 22 сентября 2014 в 16:08



Библиотеку такого типа невозможно заставить работать их коробки для абсолютно всех случаев. Есть распознавание гугла — оно работает без настроек. Но вы ничего не контролируете, кроме языка. В конфигурации rocketsphinx есть опция -rawlogdir, она задаёт путь к директории, куда сохраняются записи во время распознавания. Прослушайте записи, возможно, вы записываете не в том формате. Это очень распространённая ош Если в формате всё хорошо, то вы можете написать в список рассылки, форум или #ctusphinx на freenode, указав ссылки на записи и вашу конфигурацию. Также, сейчас есть некоторые проблемы с версионностью. По сути, сейчас можно скачать trunk/master или 0.8, и они очень сильно различаются. Я советую скачивать первую версию. Есть не собирается — пишите, такое бывает. Разработчиков немного, но все откликаются оперативно.



2d0x 22 сентября 2014 в 19:07



Спасибо за статью. Я. правильно понимаю, что можно совместить оба метода и после «условных 3 секунд», если команда не распознана, отправлять запись на серверный движок??



morfeusys 22 сентября 2014 в 19:45



Ну теоретически можно. Но на какой движок и зачем вы будете это отправлять? Тем более если вы что-то сказали, то rocketsphinx все равно вери вам какую-то гипотезу по грамматике.



ClusterM 24 сентября 2014 в 15:02



Спасибо, появился повод разобраться уже с этой библиотекой. Давно пора прикрутить её к моей Вике.



morfeusys 24 сентября 2014 в 16:08



Да, действительно. Я-то уж давно прикрутил ее к моей [Дусе](#) ;-)
[и в умный дом еще](#)



past 25 сентября 2014 в 14:35



На мой згляд, было бы здорово добавить универсализма, например отправлять результаты через mqtt.



morfeusys 25 сентября 2014 в 14:41




[Максимальный универсализм с помощью голосового ассистента и сервера openhab](#)

 **past** 25 сентября 2014 в 14:49 # 📌 🔍 ↻

К сожалению видео демонстрация это совсем не «how to».

 **morfeusys** 25 сентября 2014 в 15:03 # 📌 🔍 ↻

Вот вам how to от того же автора

 **past** 25 сентября 2014 в 15:02 # 📌 🔍 ↻

Кроме того, не OpenHAB ом единым...

 **past** 26 сентября 2014 в 08:52 # 📌 🔍 ↻

Посмотрите все же на протокол mqtt. С его помощью можно с десятками систем интегрироваться. От MySensors до node-red.


 **morfeusys** 26 сентября 2014 в 09:15 # 📌 🔍 ↻

Еще раз отвечу вам — openhab. Он интегрирован и с mqtt тоже. Openhab интегрируется с сотней систем, а не с десятком.

 **past** 26 сентября 2014 в 14:22 # 📌 🔍 ↻

Вы имеете ввиду протоколы взаимодействия с оборудованием(z-wave, x10, knx ...), а я про системы домашней автоматизации, такие как openhab, mysensors.


То есть, написав модуль, взаимодействующий с миром посредством mqtt вы потом без труда можете его подключить к любому продукту с поддержкой mqtt.

 **morfeusys** 26 сентября 2014 в 14:45 # 📌 🔍 ↻


Тогда, возвращаясь к вашей реплике

На мой взгляд, было бы здорово добавить универсализма, например отправлять результаты через mqtt.


Какие результаты вы бы предложили отправлять через mqtt???

 **past** 26 сентября 2014 в 22:42 # 📌 🔍 ↻

Номер слова/выражения из списка(как работает EasyVR) или само слово/выражение.

 **past** 26 сентября 2014 в 23:14 # 📌 🔍 ↻

Если «на той стороне» достаточно умное приложение, то всю логику, модель языка можно там выстроить.

 **morfeusys** 27 сентября 2014 в 20:27 # 📌 🔍 ↻

Простите, а зачем логика должна быть где-то?? Кокретно в данном приложении логика на клиенте и этого хватает для решения конкретной задачи...

 **past** 10 октября 2014 в 17:20 # 📌 🔍 ↻

У меня в первом приближении получилось так:

Дуся распознает Event taskера. Taskер с помощью плагина MQTT Sender отправляет сообщение с топиком, «actions/команда» и телом {id: номер}

Поток в Node-RED слушает соответствующий топик и посылает команду устройству с соответствующим айди.

 **ClusterM** 8 октября 2014 в 13:10 # 📌

Вчера наконец-то дошли руки разобраться в этой библиотеке. Сейчас я для умного дома использую онлайнную распознавалку от Google. Надо сказать у Pocketsphinx есть как плюсы, так и минусы.

Минусы:

- В «grammar-based» режиме фразы распознаются неплохо за счёт того, что используется заранее заданный список, но таким образом иногда и шум принимается за какую-то команду.
- Чтобы не было вышеописанной проблемы, приходится использовать ключевое слово для активации. Как уже сказал автор, оно не должно быть коротким. Это в разы понижает удобство использования, ведь я как-то уже привык, что ночью достаточно сказать просто «открой окно» или «включи компьютер» без всяких прелюдий в виде «умный дом». При этом моё привычное необязательное обращение по имени «Вика» слишком коротко.
- Всё-таки плохо распознаётся с большого расстояния. Или я что-то плохо настроил? Но пока я тестировал Pocketsphinx, на мои команды гораздо реагировал планшет, который находится в другом конце комнаты и использует распознавание от Google.

Плюсы:

- Скорость. Абсолютно мгновенная реакция на команды. В случае с Google приходится ждать лишнюю секунду, а если использовать Android 4.4, то за костылей иногда доходит и до пяти секунд.
- Автономность, не нужен Интернет.

Пока что минусов для меня больше :(



morfeusys 8 октября 2014 в 14:32



Первая проблема — не проблема. Решается за счет правильно написанной грамматики (garbage) и фильтрации результат по score, который можно легко посчитать. Это обычная практика для распознавания на базе контекстно-свободной грамматики.

Второе — «в разы понижает удобство»??? Я как-то не заметил этого «в разы». А вы гуглом все время слушаете эфир на предмет команды?? Ого вас забанит или весь трафик съест :) Плюс гугл любит становиться «недоступным» — это все же облачный сервис...

Третье — еще все зависит от микрофона. Вы говорите что гугл распознает с нескольких метров?? Не поверю. Гугл распознавание рассчитано на небольшое расстояние. А в шуме вы от него вообще не добьетесь результата на расстоянии метра.



ClusterM 8 октября 2014 в 14:44



Вот пишу файл грамматики, действительно становится лучше.

Да, постоянно слушаю эфир, но при этом из общего шума вполне неплохо выделяется голос, так что запросы происходят не непрерывно. Никаких проблем.

В почти полной тишине (шум компьютера) — Google легко распознаёт с нескольких метров. В абсолютно полной тишине берёт и с пяти-семи метров. Правда, так было не всегда. Впечатление, что в последнее время они как-то оптимизировали распознавание. Меня это даже напугало немного. Pocketsphinx шум тоже как-то не очень любит



morfeusys 8 октября 2014 в 15:07



Шум не любит не покетсфинкс, а я конкретная акустическая модель. Если натренировать свою — будет работать и в шуме.

Если постоянно слушать эфир... Ну что тут сказать — трафик плюс гугл в конечном итоге откажет вашему приложению в обслуживании, как только вашей системой начнете пользоваться не только вы...



ClusterM 8 октября 2014 в 15:21



У API ключей Гугла вполне себе определённые ограничения по запросам в сутки, за пределы которых я не выхожу.



morfeusys 8 октября 2014 в 18:25



Ну это только вы :) и сколько же у них ограничения? :) вообще-то под андроидом ограничения нет, но все же — слушать постоянно гоглабе... это как-то что-то странно... И гугл будет очень недоволен, если это будет делать еще хотя бы пара тысяч человек...



BelBES 9 октября 2014 в 14:07



Называй теперь свой дом вместо Вики какой-нибудь Алефтиной и двух проблем сразу не станет)

А проблему с дистанцией, как мне кажется, можно решить использованием нескольких микрофонов по всей квартире с равномерным покрытием площади. От этого все равно никуда не денешься, разве что на себе микрофон носить.



yvm 19 декабря 2014 в 13:18



«Окей, Вика» || «Короче, Вика»



Jey 26 января 2015 в 19:16



Добрый день!

Подскажите пожалуйста такой вопрос по приведённым исходникам, как поменять ключевую фразу активации?

Нашёл, что она прописана в strings.xml как «Умный дом», так же в файле hotwords как два слова:

умный ui m n ay j

дом d oo m

пытался менять на свою фразу, но что-то ничего не получилось. нужно ли менять её в обоих местах (в каком-то особом формате?) или как-то ещё другому?

P.S. может кому-нибудь пригодится — у меня не получалось запустить движок без добавления двух файлов-библиотек в проект:

libs\armeabi\libpocketsphinx_jni.so

libs\armeabi-v7a\libpocketsphinx_jni.so




raid 10 февраля 2015 в 10:09



Если в этой статье ещё найдутся живые люди, помогите понять, что есть score?

**Guzzle** 10 февраля 2015 в 16:11    

Рискуны предположить, что score это коэффициент схожести распознанного с активационной фразой.

**raid** 10 февраля 2015 в 16:58    





Это я тоже рискнул предположить.

Score, кстати, указывается не только для активационной фразы, но и для любой распознанной (включая распознанную по грамматике).

Вот только как с ним правильно работать и от чего конкретно он зависит, это уже более сложный вопрос.

**nix0iD** 24 октября 2015 в 11:12  

А где реализация getAssets()?

**nix0iD** 24 октября 2015 в 11:17    

Прошу прощения, разобрался, первый проект под Android)

**Ghool** 16 января 2016 в 00:00  

А что с офлайн-распознаванием под Windows?

Был Realspreeker, но сейчас не работает, обещают новую версию.

Да и работал он онлайн, судя по тому, что сейчас не действуют старые дистрибутивы.

Был когда-то горыныч, но давно он был так себе.

Что рекомендуется сейчас для набора текстов?

**Bangybug** 23 марта 2016 в 10:59    

библиотека собирается под большинство архитектур и операционных систем, windows в их числе

**Bangybug** 23 марта 2016 в 07:08  

Совсем никак он не распознаёт мой английский, даже если я говорю те же фразы что в демо-файлах (их он распознаёт). При этом я уверен в том, что формат моих записей правильный. В общем, сразу как-то не заладилось.

**Bangybug**  23 марта 2016 в 10:58    

Сегодня мне ответили на форуме, причина была в том, что алгоритм не успевает адаптироваться к распознаванию коротких фраз. Эту адаптацию провести самому, что не должно быть ни долго, ни сложно — по идее, около минуты поговорить надо, извлечь получившиеся коэффициенты и поместить их в конфигурацию.

**Afganec** 23 марта 2016 в 11:44    

Спасибо вам! Очень ценное дополнение, у меня тоже были подобные проблемы. Пойду достану проект из шкафчика и попробую еще раз.

Только полноправные пользователи могут оставлять комментарии. Войдите, пожалуйста.

САМОЕ ЧИТАЕМОЕ

Сутки

Неделя

Месяц

В Госдуму внесен законопроект об автономной работе рунета

 +63

 36k

 50

 388

Вопросы не мальчика, а джуна. 22 вопроса работодателю на собеседовании на позицию «Middle Python-разработчик»

 +31

 23,1k

 223

 93

Повсеместное размытие спутниковых фотографий раскрывает местонахождение секретных баз

+30

15,8k

28

56

Иди-ка ты сам на... или правила общения в команде

+176

39,3k

114

255



В гости к отцу

+48

107k

55

58

Аккаунт	Разделы	Информация	Услуги	Приложения
Войти	Публикации	Правила	Реклама	<div></div>
Регистрация	Хабы	Помощь	Тарифы	
	Компании	Документация	Контент	
	Пользователи	Соглашение	Семинары	
	Песочница	Конфиденциальность		