Публикации

Пользователи

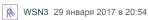
Хабы

Компании

Песочница

Войти

Регис



# Джарвис снова в деле

DIY или Сделай сам

Из песочницы

Наверняка, каждый мечтает о своем голосовом ассистенте, под катом еще одна реализация «Джарвиса» из известного фильма.



Давно не покидала мысль о своем «Jarvis» и управлении техникой в доме голосом. И вот, наконец, руки дошли до создания сего чуда. Над «мозгами» долго думать не пришлось, Raspberry Pi подходит идеально.

Итак, железо:

- Raspberry pi 3 model b
- USB камера logitech

# Реализация

Работать наш ассистент будет по принципу Alexa/Hub:

- 1. Активироваться оффлайн по определенному слову
- 2. Распознать команду в облаке
- 3. Выполнить команду
- 4. Отчитаться о проделай работе либо сообщить запрашиваем информацию

Т.к. моя камера поддерживается из коробки, с драйверами возиться не пришлось, поэтому сразу переходим к программной части.

## Оффлайн активация

Активация будет происходить с помощью CMU Sphinx, и все бы хорошо, но из коробки распознание происходит очень медленно, больше 1 что абсолютно не подходит, для решения проблемы нужно очистить словарь от ненужных слов.

Устанавливаем все необходимое:

```
pip3 install SpeechRecognition
pip3 install pocketsphinx
```

#### далее

```
sudo nano /usr/local/lib/python3.4/dist-packages/speech_recognition/pocketsphinx-data/en-US
/pronounciation-dictionary.dict
```

удаляем все кроме нужного нам Джарвиса:

```
jarvis JH AA R V AH S
```

Теперь pocketsphinx распознает довольно быстро.

### Распознавание речи

Сначала была идея воспользоваться сервисом гугла, к тому же его поддержка есть в SpeechRecognition. Но как оказалось гугл берет за этс деньги и не работает с физ. лицами.

Благо Яндекс тоже предоставляет такую возможность, бесплатно и предельно просто.

Регистрируемся, получаем API KEY. Все работу можно производить curl'om.

```
curl -X POST -H "Content-Type: audio/x-wav" --data-binary "@file" «https://asr.yandex.net/asr_xml?uuid=ya_uid&key=yf_api_k topic=queries»
```

# Синтез речи

Тут нам опять поможет Яндекс. Посылаем текст в ответ получаем файл с синтезированным текстом

```
curl whttps://tts.voicetech.yandex.net/generate?format=wav&lang=ru-RU&speaker=zahar&emotion=good&key=ya_api_key» -G --data lencode "text=text" > file
```

#### **Iarvis**

Собираем все вместе и получаем такой скрипт.

```
#!/usr/bin/env python
# -*-coding:utf-8-*-
import os
import speech_recognition as sr
from xml.dom import minidom
import sys
import random

r = sr.Recognizer()
ya_uuid = ''
ya_api_key = ''

# os.system('echo "Accucr+ehr san+ymeh" | festival --tts --language russian')

def convert_ya_asr_to_key():
    xmldoc = minidom.parse('./asr_answer.xml')
    itemlist = xmldoc.getElementsByTagName('variant')
    if len(itemlist) > 0:
```

```
return itemlist[0].firstChild.nodeValue
   else:
       return False
def jarvis on():
   with sr.WavFile("send.wav") as source:
       audio = r.record(source)
       t = r.recognize_sphinx(audio)
       print(t)
   except LookupError:
       print("Could not understand audio")
   return t == ("jarvis")
def jarvis_say(phrase):
   os.system(
        'curl "https://tts.voicetech.yandex.net/generate?format=wav&lang=ru-RU&speaker=zahar&emotion=good&key='+ya_api_key
-G --data-urlencode "text=' + phrase + '" > jarvis_speech.wav')
   os.system('aplay jarvis_speech.wav')
def jarvis_say_good():
   phrases = ["Готово", "Сделано", "Слушаюсь", "Есть", "Что-то еще?", ]
   randitem = random.choice(phrases)
   jarvis_say(randitem)
try:
   while True:
       os.system('arecord -B --buffer-time=1000000 -f dat -r 16000 -d 3 -D plughw:1,0 send.wav')
       if jarvis_on():
           os.system('aplay jarvis_on.wav')
           os.system('arecord -B --buffer-time=1000000 -f dat -r 16000 -d 3 -D plughw:1,0 send.wav')
                'curl -X POST -H "Content-Type: audio/x-wav" --data-binary "@send.wav" "https://asr.yandex.net/asr_xml?uui
+ya uuid+'&key='+ya api key+'&topic=queries" > asr answer.xml')
            command_key = convert_ya_asr_to_key()
            if (command key):
               if (command_key in ['key_word', 'key_word1', 'key_word2']):
                   os.svstem('')
                   jarvis_say_good()
                    continue
except Exception:
jarvis_say('Что-то пошло не так')
```

Что тут происходит. Запускаем бесконечный цикл, arecord'om записываем три секунды и отправляем sphinx на распознание, если в файле встречается слово «jarvis»

```
if jarvis_on():
```

проигрываем заранее записанный фаил оповещения об активации.

Опять записываем 3 секунды и отправляем Яндексу, в ответ получаем нашу команду. Далее выполняем действия исходя из команды.

На этом собственно все. Сценариев выполнения можно придумать великое множество.

#### Use-case

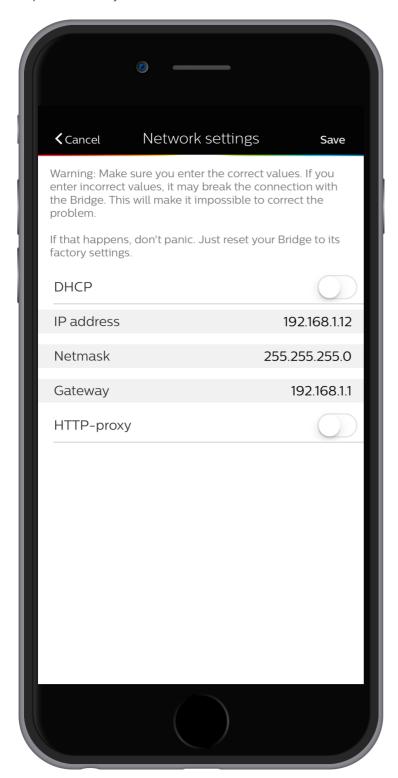
Теперь немного примеров реального моего использования

# **Philips Hue**

#### Устанавливаем

pip install phue

В приложении Hue устанавливаем статический IP:



## Запускаем:

#!/usr/bin/python
import sys
from phue import Bridge

```
b = Bridge('192.168.0.100') # Enter bridge IP here.

#If running for the first time, press button on bridge and run with b.connect() uncommented
#b.connect()

print (b.get_scene())
```

Выписываем ID нужных схем, вида «470d4c3c8-on-0»

#### Конечный вариант скрипта:

```
#!/usr/bin/python
import sys
from phue import Bridge

b = Bridge('192.168.0.100') # Enter bridge IP here.

#If running for the first time, press button on bridge and run with b.connect() uncommented
#b.connect()

if (sys.argv[1] == 'off'):
    b.set_light([1,2,3],'on', False)
else:
    b.activate_scene(1,sys.argv[1])
```

#### В джарвиса добавляем:

```
if (command_key in ['включи свет', 'включить свет', 'свет']):
    os.system('python3 /home/pi/smarthome/hue/hue.py all67aa91-on-0')
    jarvis_say_good()
    continue

if (command_key in ['приглуши свет', 'приглушить свет']):
    os.system('python3 /home/pi/smarthome/hue.py ac637e2f0-on-0')
    jarvis_say_good()
    continue

if (command_key in ['выключи свет', 'выключить свет']):
    os.system('python3 /home/pi/smarthome/hue/hue.py "off"')
    jarvis_say_good()
    continue
```

# **LG TV**

Берем скрипт отсюда. После первого запуска и ввода кода сопряжения, сам код не меняется, поэтому можно выпилить эту часть из скрипта оставить только управляющую.

В джарвиса добавляем:

```
#1 - POWER

#24 - VOLUNE_UP

#25 - VOLUME_DOWN

#400 - 3D_VIDEO

if (command_key in ['выключи телевизор', 'выключить телевизор']):

os.system('python3 /home/pi/smarthome/TV/tv2.py 1')

jarvis_say_good()

continue

if (command_key in ['прибавь громкость', 'громче']):

os.system('python3 /home/pi/smarthome/TV/tv2.py 24')

jarvis_say_good()

continue
```

#### Радио

```
sudo apt-get install mpg123
```

#### В джарвиса добавляем:

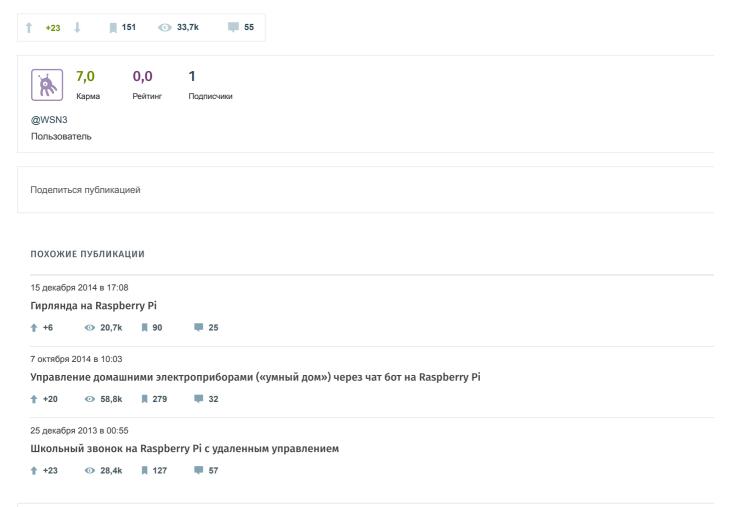
```
if (command_key in ['новости', 'выключи новости','что происходит']):
os.system('mpg123 URL')
continue
```

Еще можно поставить homebridge и управлять всем через Siri, в случае если до джарвиса не докричаться.

Что касается качества распознавания речи, не Alexa конечно, но на расстоянии до 5 метров процент верного попадания приличный. Главна проблема — речь из телевизора\колонок записывается вместе с командами и мешает распознаванию.

На этом все, спасибо.

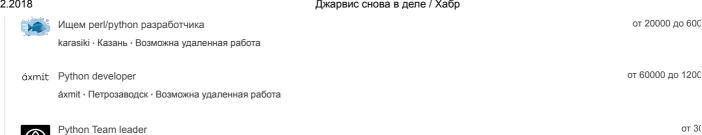
**Теги:** raspberry pi, python





Программист Python Хантфлоу · Махачкала · Возможна удаленная работа

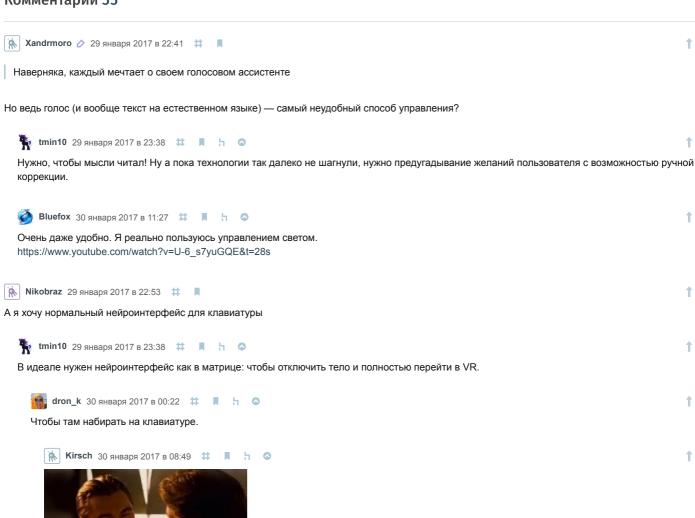
от 10000 до 200



Все вакансии

GetPure Inc · Возможна удаленная работа

# Комментарии 55





Пока текущие возможности и технологии по набору текста при помощи нейроинтерфейсов сильно уступают в скорости набора.

kernel24 30 января 2017 в 00:20 #

Чего уже мелочиться, зачем вам это бренное тело, всех в матрицу :)

🗼 **Jogger** 30 января 2017 в 00:57 👯

>Наверняка, каждый мечтает о своем голосовом ассистенте

Только если он будет комплектоваться телом-андроидом. Я совершенно не представляю, зачем мне голосовой помошник, который не может ничего

сделать. Большинством функций умного дома, как по мне, куда удобнее рулить с графического или текстового интерфейса. Вот попробовал я Корта ну, поигрался немного, но так и не придумал ей практического применения. Может быть об этом и мечтают люди, у которых нет навыка работы с клавиатурой (хотя опять же, в большинстве случаев хватает навыков работы с мышью или тач-панелью), но зачем такой помощник гикам — ума не приложу.



Не всегда, например, мне просто лень что-то тыкать на планшете (на планшете это будет несколько нажатий, проще уж воспользоваться обычны выключателем), когда я захожу домой, проще голосом сказать свет. И далее остальные применения, по той же схеме. Можно взглянуть на это инк кому как удобнее.



Опять же, мне проще щёлкнуть выключателем, чем что-то вслух произносить. И намного быстрее. И шансов что мою команду неправильно распознают чуток поменьше. И ложных срабатываний почти нет (разве что если спиной заденешь). Нет, ну пожалуй если у кого-то квартира большая, а выключатели расположены неудобно, и зайдя в комнату до выключателя надо идти — может и удобно, я ж не знаю. Но как-то слиц много условий чтобы писать «каждый мечтает».



Удобно активировать различные сценарии. Например: Кино, музыка, уборка и т.д. где активируется ряд действий, например чтобы у меня рувключить режим «кино». то нужно взять 3 пульта, от телевизора, ресивера и проектора, плюс встать и выключить везде свет.



Я Алексу прошу

- \* Включить музыку пока готовлю
- \* Включить/выключить свет на кухне если руки заняты
- \* Включить/выключить монитор (на нем нет физической кнопки)

🦍 gsaw 30 января 2017 в 11:50 # 📕 🤚 🔕

- \* Включить/выключить свет в зале когда мы на диване телек смотрим к примеру
- \* Поставить таймер на N минут
- \* Узнать погоду
- \* Включить/выключить группу приборов
- \* Включить отопление

(del)

В самом начале еще всякую ерунду типа «сколько лет Путину» или «какое расстояние до солнца» спрашивали. Но быстро надоело. Так да, выключатель в пределах пары метров, то проще им щелкнуть, чем выговаривать фразу, которую надо еще правильно сформулировать. Но Echo dot стоит потраченных 50 Евров, хотя и нервирует иногда, своей тупостью.



нет конечно, к сожалению русского нет и я думаю никода не будет. Вся надежда на google home. Google Assistent уже понимает русски письменный, а голосовое управление еще нет.





Включить/выключить свет в зале когда мы на диване телек смотрим к примеру

А как успехи с распознаванием голоса при работе какого-либо фонового звука? Так же хотелось бы услышать ответ на этот вопрос от автора статьи.



Про Алексу. Если тишина, то голос не надо повышать, можно обычным тоном говорить. Если в соседней комнате работает телевизор, распознованию практически не мешает. Если в той же комнате (у меня это кухня) булькает к примеру кофеварка или работает вытяжк приходится повышать голос. Иначе либо не понимает, либо «делает вид», что ничего не услышала. Если телевизор сделать совсем гр то приходится громко говорить. Если алекса сама музыку воспроизводит, то достаточно, что бы она услышала свое имя, она тут же приглушает звук. У меня зал по соседству с кухней, из зала я тоже могу отдавать комманды, правда приходиться повышать голос и глу телевизор.

В принципе неплохо работает, иногда есть какие то непонятки. Некоторые слова упорно не хочет понимать, возможно потому, что я на иностранном для меня языке говорю. Хотя в логе видно, что предложение поняла. Говорю к примеру «включи монитор», она все распознала, видно в логе, но все равно переспрашивает, какой девайс я имел ввиду. Если тут скажу «монитор» она его включает. Я об эту проблему переименовав монитор в «компьютер». Стала понимать слету, даже шепотом.

Но вот из-за таких непоняток невольно начинаешь повышать голос и говорить как робот. У меня такое чуство, что проблема не в распозновании речи как таковой, а понимании семантики.

НЛО прилетело и опубликовало эту надпись здесь



По моему личному мнению, с вероятностью в 99,9% такого не будет в ближайшие не знаю сколько лет. А то, может и никогда не будет.

Поэтому только условные сценарии, которые более-менее покрывают потребности обитателей. Ну либо мириться с приличным количествого ошибок, если полагаться на некое подобие искусственного интеллекта, который пытается силой своего скудного разума пронзить привычки человека.



Ну почему. Главная проблема точного детекта людей в квартире. Если решить её (например продвинутыми датчиками движения, может какие-то примитивные 2д лидары), то можно запустить систему в режим обучения, чтобы она изучила привычки людей и потом старалас воспроизводить. Раз учат нейронки водить авто, то и угадывать, что если я захожу в комнату и там темно, то нужно включить свет уже не сложно.

 spc
 30 января 2017 в 11:05
 #
 ■
 □

И вот я снова становлюсь занудой. Собственно, здесь (на ГТ) каждый, кому не лень, говорит, что то, что мы называем умным домом, з умный дом.

Вот и я соглашусь: сейчас дома автоматические. А вот до умного нам как до Марса пешком.

Предположим, вы обклеите все датчиками и даже поставите камеры, которые будут (вместе с другим железом) распознавать пол и во каждого. Но вот ситуации:

- а) Вы приходите домой один
- б) Вы приходите домой с известной дому дамой
- в) Вы приходите с неизвестной дому дамой
- г) Вы приходите домой с известной дому дамой, которой неизвестно, что в пункте в) вы приходили домой с неизвестной дамой
- д) Вы приходите домой с известной дому дамой, с которой романтика уже закончилась

Как дом будет понимать кого как встречать, ну или хотя бы кому какой свет — где поярче, где потише? Вот поэтому я и говорю, что пок можно базироваться только на сценариях, которые более-менее учитывают общие алгоритмы поведения.



А зачем прямо идеально чтобы реагировал? Он должен кпредугадывать что-то несложное, остальное можно и руками подкрутить, это довольно редко. Т.е. грубо говоря работать на простых кейсах, которые покроют 90% времени: включить свет по приходу, сопровождать движение по коридорам их подсветкой, вечером закрыть жалюзи, утром их открыть, согласуясь с будильником.

 spc 30 января 2017 в 11:49
 #
 Д

 ф
 В
 В
 В

 ф
 В
 В
 В

 ф
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

 В
 В
 В
 В

Повторюсь, там вон выше пишут «любой умный дом». Вот именно поэтому.

НЛО прилетело и опубликовало эту надпись здесь



Я вижу вы с новой спутницей?

Вот это как раз глупый дом )

Отчего же глупый? Просто говорить это надо по нейроинтерфейсу или, хотя бы, микронаушнику :)



PIR датчики влепили и работает)

НЛО прилетело и опубликовало эту надпись здесь



В принципе, ничего не мешает сделать Джарвиса для обычного ноутбука.

Ты за ноутом, микрофон на расстоянии максимум метра.

Команды типа «Включи новости/музыку», «зайди на Гиктаймс/Баш/ещё-куда-нибудь».

imm 30 января 2017 в 08:36 # ■

Команда «б\*\*ь» отменяет последнее действие?

Mallexin 30 января 2017 в 10:01 # ■

Я вот чего не понимаю:

телевизор, музыкальая система, вентиляция — все эти штуки являются частью системы умного дома.

Почему нигде не внедряют систему распознавания голоса для умного дома, которая из звукового потока убирает известыне ей звуки?

Звук из телевизора мешает распознавания? Да нет проблем, просто уберите из входного потока звук телевизора. У вас же есть доступ к звуковому п телевизора.

Truin10 30 января 2017 в 10:50 # Л о

Звук-то не идеально будет идти в микрофоны, стоит учитывать переотражения от стен, поглощение и прочее.

НЛО прилетело и опубликовало эту надпись здесь

Инсендио!!! В сторону люстры )

Нет уж, надо чтобы распознавание было хорошим. Мы же легко распознаем какой угодно голос и компьютер так сможет, если уже не может.

НЛО прилетело и опубликовало эту надпись здесь

Я вот думал — пульт ду на e-lnk, с гироскопом и может быть с компасом. К примеру настроить несколько схем. Для выключения света две кнопки телевизора много специфичных кнопок, для управления жалюзями тоже две кнопки и может быть контрол для диммера. А потом выбрать схему *д* лампочки, направить пульт на лампочку и зафиксировать положение гироскопа и компаса. Так же с телевизором и окном. Потом направил на теле пульт сам включит схему с кнопками для телефизора. Направил вверх на лампочку и автоматически активируется схема для управления лампой.

Конечно это будет работать, если пульт всегда только на одном месте будет использоваться. У меня по крайней мере так. Грубо направления лев окно, прямо — телевизор, вверх — лампа должно быть достаточно для выбора лайоута.

НЛО прилетело и опубликовало эту надпись здесь

Надо сделать звуковую отвертку!

Если отслеживать в пространстве, то в любом случае потребуется система определения куда направлено, с метками пространства какими-Чтоб понятно было, что вот заклинание «светиус-медиус» вы на люстру центральную кастанули, а не на бра.

Так что было бы веселее «Джарвис», рукой на люстру, по часовой покрутили — свет прибавился. Тоже самое — на кондиционер и телевизо Без «пульта» или палочек, чтоб каждый раз эти костыли не искать.

А для этого нужна система распознавания жестов независимо от освещенности.

НЛО прилетело и опубликовало эту надпись здесь

и камерой(камерами) определяем положение и направление.

ИК камерами?

Или в темноте волшебство не случится?

Плюс необходимо уверенное опознавание во всех углах квартиры (т.е. количество камер большое). И привязать палочку-пульт к себе.

НЛО прилетело и опубликовало эту надпись здесь

1. Еще раз: привязанная к вам палочка с гироскопом и кнопкой.

Вы пульт телевизора никогда не теряли? Не приходилось его находить в холодильнике?

Плюс, как точно гироскоп поможет определить, что вы сейчас показываете на левее севера на 2 градуса, а не 3? Тогда не забудьте добавить в нее звуковой сигнализатор на команду «джарвискудамлятьделасьэтачертовапалочка»

2) Куда мигают? Сколько приемников надо ставить для каждой комнаты для уверенного приема везде? Данные через ик моргани точно передавать не стоит.

НЛО прилетело и опубликовало эту надпись здесь



Так это, вайфай же для определения движений и вперед! )))

**10 noonv** 30 января 2017 в 11:05

Расскажите подробнее про качество распознавания. Какой процент ложных срабатываний?



🍘 Bluefox 30 января 2017 в 11:30 💢 📗

Вот бы видео посмотреть, как это работает. Если тормоза (дольше 2 секунд), то пользователю надоест ждать.

kazenniy 30 января 2017 в 12:25 # 📕

Яндекс предоставляет бесплатно говорите? А в условиях использования SpeechKit Cloud пишут: «Для знакомства с технологией предусмотрен бесплатный тестовый период — 1 месяц с момента отправки первого запроса на сервер.

Чтобы продолжить использовать SpeechKit Cloud после этого, необходимо заключить договор. Можно приобрести пакет (фиксированное количество запросов в месяц) или оплачивать запросы по факту.

Стоимость лицензии зависит от количества обращений и в среднем составляет 400 рублей за 1000 запросов.»

Yandex SpeechKit Cloud — платный продукт. Тем не менее мы даём возможность использовать его бесплатно, если Ваш проект подходит под одн этих категорий:

- для личного использования (система «Умный дом», персональный ассистент, робот, голосовое управление ПК);
- образовательный (для подготовки диплома, реферата или обучения студентов);
- благотворительный (направленный на помощь людям с ограниченными возможностями).

При этом Ваш сервис должен отправлять к серверу Яндекса не более 1000 запросов в сутки. Если Вы рассчитываете, что запросов будет больше сообщите нам об этом на voice@support.yandex.ru.

Segmentq 30 января 2017 в 15:21 # 📕 🦙 🔕

То-то я думаю, в воздухе витает устойчивый запах яндекса, и действительно — реклама собственной персоной:)

А вообще была тут как-то статься аналогичная только с сервисом гугла.

🊺 jhonyxakep 30 января 2017 в 19:33 🗰 📙

Как вариант использовать встроенный в Google Chrome и Chromium голосовой движок. Он работает локально и умеет в непрерывное распознавание минусов конечно сам Chrome, он тяжелый для Малинки.

Для полноценного использования можно реализовать с помощью Electron, там тоже работает распознавание

🤗 **nlx0iD** 17 января 2018 в 13:01 💢

Смущает только ограничение в 3 секунды. Я делал для android-а, там он слушает, пока продолжаешь говорить. Уже год прошел — никак не дорабат этот момент?

пlx0iD 19 января 2018 в 14:07 #

Ошибка в коде кстати:

```
'arecord -B --buffer-time=1000000 -f dat -r 16000 -d 3 -D plughw:1,0 send.wav'
```

Либо -В лишнее, либо --buffer-time

Только полноправные пользователи могут оставлять комментарии. Войдите, пожалуйста

