Transcription on the Cloud

An introduction from Morgan Strong

Access updates

WCAG 2.1 has reached candidate level.

What does this mean?

Unless real world testing reveals issues, it'll be verified. Expect a mid-2018 release.

Resources: Web policies

Web Accessibility Laws and Policies by country - https://www.w3.org/WAI/Policy/

What is this about?

Definitive list of the relevant laws and policies in place in many countries around the world.

Resources: IoT Accessibility

Web of things - https://www.w3.org/WAI/APA/wiki/Web_of_Things

What is this about?

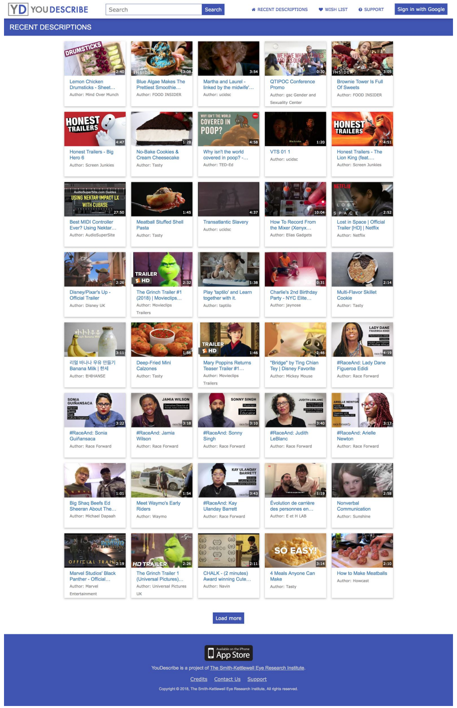
Chance to capture use cases and resources relating to accessibility of Internet of Things (IoT) through the Web of Things (WoT)

New tech to keep an eye on

YouDescribe - https://youdescribe.org/

What is it?

“Sighted people view YouTube videos and record descriptions of what they see. When the video is played with YouDescribe, the descriptions are played back with the video.”



RECENT DESCRIPTIONS

Lemon Chicken Drumsticks - Sheet... Author: Mind Over Munch

Blue Algae Makes The Prettiest Smoothie... Author: FOOD INSIDER

Martha and Laurel - linked by the midwife'... Author: ucidsc

QTIPOC Conference Promo Author: gsc Gender and Sexuality Center

Brownie Tower Is Full Of Sweets Author: FOOD INSIDER

Honest Trailers - Big Hero 6 Author: Screen Junkies

No-Bake Cookies & Cream Cheesecake Author: Tasty

Why isn't the world covered in poop? -... Author: TED-Ed

VTS 01 1 Author: ucidsc

Honest Trailers - The Lion King (feat.... Author: Screen Junkies

Best MIDI Controller Ever? Using Nektar... Author: AudioSuperSite

Meatball Stuffed Shell Pasta Author: Tasty

Transatlantic Slavery Author: ucidsc

How To Record From the Mixer (Xenyx... Author: Elias Gadgets

Lost in Space | Official Trailer [HD] | Netflix Author: Netflix

Disney/Pixar's Up - Official Trailer Author: Disney UK

The Grinch Trailer #1 (2018) | Movieclips... Author: Movieclips Trailers

Play 'taptilo' and Learn together with it. Author: taptilo

Charlie's 2nd Birthday Party - NYC Elite... Author: jaynose

Multi-Flavor Skillet Cookie Author: Tasty

리얼 바나나 우유 만들기 Banana Milk | 한세 Author: 한세HANSE

Deep-Fried Mini Calzones Author: Tasty

Mary Poppins Returns Teaser Trailer #1... Author: Movieclips Trailers

"Bridge" by Ting Chian Tey | Disney Favorite Author: Mickey Mouse

#RaceAnd: Lady Dane Figueroa Edidi Author: Race Forward

#RaceAnd: Sonia Guiñansaca Author: Race Forward

#RaceAnd: Jamia Wilson Author: Race Forward

#RaceAnd: Sonny Singh Author: Race Forward

#RaceAnd: Judith LeBlanc Author: Race Forward

#RaceAnd: Arielle Newton Author: Race Forward

Big Shaq Beefs Ed Sheeran About The... Author: Michael Dapaah

Meet Waymo's Early Riders Author: Waymo

#RaceAnd: Kay Ulanday Barrett Author: Race Forward

Évolution de carrière des personnes en... Author: E et H LAB

Nonverbal Communication Author: Sunshine

Marvel Studios' Black Panther - Official... Author: Marvel Entertainment

The Grinch Trailer 1 (Universal Pictures)... Author: Universal Pictures UK

CHALK - (2 minutes) Award winning Cute... Author: Navin

4 Meals Anyone Can Make Author: Tasty

How to Make Meatballs Author: Howcast

YouDescribe is a project of The Smith-Kettlewell Eye Research Institute.

Credits Contact Us Support

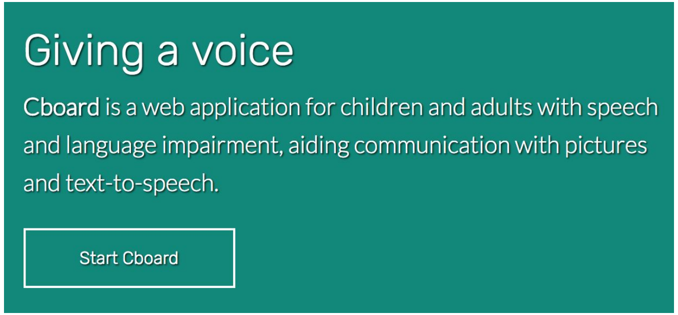
Copyright © 2018, The Smith-Kettlewell Eye Research Institute, All rights reserved.

New tech to keep an eye on

Cboard - https://www.cboard.io/

What is it?

Application that uses a board of icons and images to convert images to speech in multiple languages.



Giving a voice

Cboard is a web application for children and adults with speech and language impairment, aiding communication with pictures and text-to-speech.

Today’s topic

Using cloud based automated transcription services

How does it work?

Essentially, you put up audio onto a cloud based service, it processes the audio track using some speech recognition algorithms and produces text.

What are these services?

Amazon has released *Amazon Transcribe*
https://aws.amazon.com/transcribe/

Google Cloud has a *Cloud Speech-to-Text* service
https://cloud.google.com/speech-to-text/
powered by the Web Speech API
https://w3c.github.io/speech-api/speechapi.html

And others… Azure has Cognitive Services, and IBM has Watson Speech to text… but I focused on Amazon for this talk.

SaaS providers

Deepgram: https://deepgram.com/

Voicebase: https://www.voicebase.com/speech-to-text/

Trint: https://trint.com/ (Hybrid automatic / human)

Free online & desktop apps

Now Transcribe: http://www.nowtranscribe.com/ (software in beta)
Speechlogger: https://speechlogger.appspot.com/en/ (web - awesome)

Roughly 49 APIs here:
https://www.programmableweb.com/search/speech%20recognition

Even one for for piano:
Automatic Music Transcription with Deep Neural Networks:
https://github.com/jsleep/wav2mid

Recent open sourcing-ness

Deepgram has open sourced their materials:

Read blog here: https://techcrunch.com/2017/10/10/deepgram-opens-up-its-machine-transcription-platform-to-everyone/

Wavenet: https://deepmind.com/blog/wavenet-generative-model-raw-audio/

For those worried about the future, worry some more with this available:
Lyrebird: https://lyrebird.ai/

What are doing today?

Amazon Transcribe
https://aws.amazon.com/transcribe/

Also, taking a look at Amazon Rekognition:
https://aws.amazon.com/rekognition/

It’s not related to transcription, but for image recognition and classification. It’s not intended as an OCR replacement...

What are doing today?

We’ll also live demo:

The Web Speech API
https://www.google.com/intl/en/chrome/demos/speech.html

Speechlogger:
https://speechlogger.appspot.com/en/

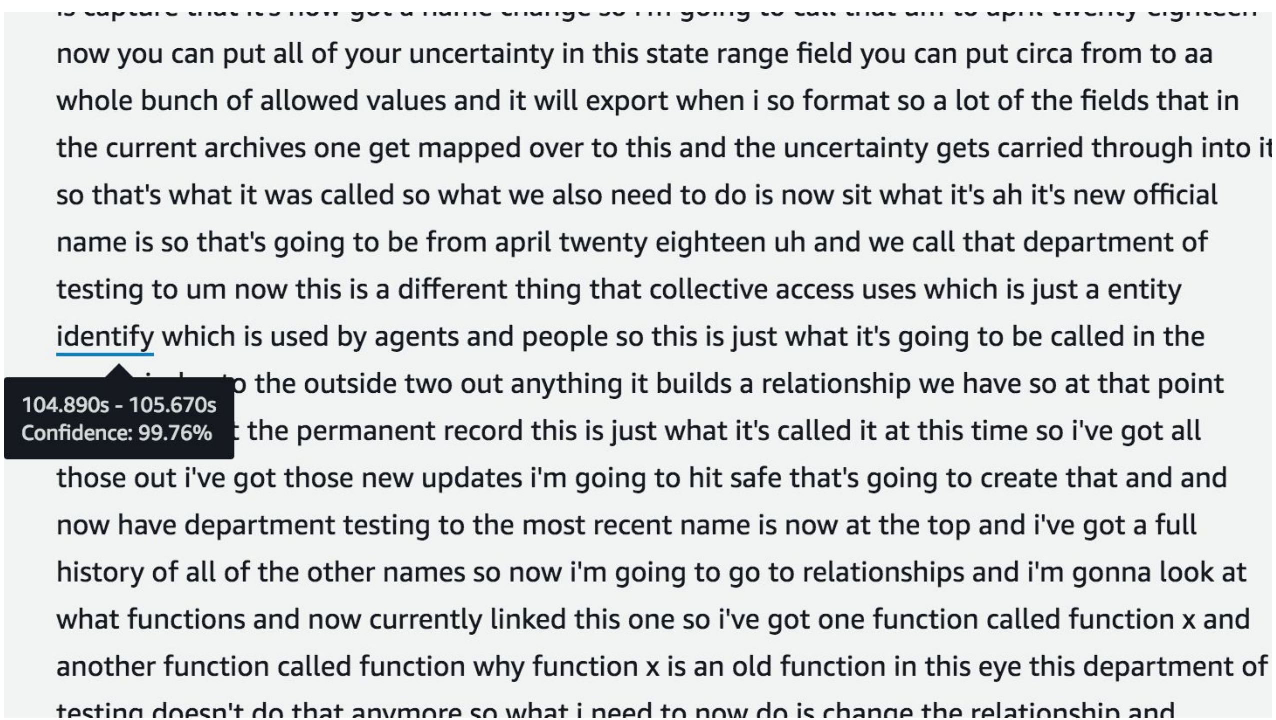
But before we get too excited...

None of is even close to perfect yet.

“ASR is not solved,” Scott Stephenson, co-founder and CEO of Deepgram, explained to me in an interview. “It’s solved for specific data sets but with noisy accented call data, any service will do a poor job with it.”

Interview with John Mannes Oct 2017.

But before we get too excited...



is capture that it's now got a name change so i'm going to call that um to april twenty eighteen
now you can put all of your uncertainty in this state range field you can put circa from to aa
whole bunch of allowed values and it will export when i so format so a lot of the fields that in
the current archives one get mapped over to this and the uncertainty gets carried through into it
so that's what it was called so what we also need to do is now sit what it's ah it's new official
name is so that's going to be from april twenty eighteen uh and we call that department of
testing to um now this is a different thing that collective access uses which is just a entity
identify which is used by agents and people so this is just what it's going to be called in the
[illegible] to the outside two out anything it builds a relationship we have so at that point
[illegible] the permanent record this is just what it's called it at this time so i've got all
those out i've got those new updates i'm going to hit safe that's going to create that and and
now have department testing to the most recent name is now at the top and i've got a full
history of all of the other names so now i'm going to go to relationships and i'm gonna look at
what functions and now currently linked this one so i've got one function called function x and
another function called function why function x is an old function in this eye this department of
testing doesn't do that anymore so what i need to now do is change the relationship and

104.890s - 105.670s
Confidence: 99.76%

Let’s take a look

… but before we, does everyone know about how cloud services work or about Amazon Web Services?

The test for transcription

I uploaded a 5:31 audio clip, which I made earlier in the month.

It was audio track of a screencast I made for a client to showcase.

One speaker, no background noise. Spoke at my regular pace, which is a bit faster than most people, and with my standard accent. Which is also, fairly standard for an Australian.

Also did a quick 1:03 very slowly spoken audio recording as a control.

The test for OCR

From my WA Museum days:
http://museum.wa.gov.au/research/records-supplements/records/food-resources-aborigines-south-west-western-australia

We had a request for an accessible version of this document, so I OCR’d it, then corrected the text. Wanted to compare accuracy of Rekognition with Adobe CS3. I did one page - Page 3.

DISCLAIMER!!!!

Before I started writing this talk, I had never used either service.

I skim read the documentation after putting my son to bed on Sunday night.

My test is also statistically invalid - three tests in total.

I am not an expert in either service, I am merely road testing, and reporting on the result that I experienced.

How they work

Utilising machine learning algorithms, that improve with more similar materials.

Both analyse the material, and produce a JSON log of each component (in OCR it's an X,Y value bounding box with a confidence rating; in transcription it's a time range as the bounding box with a confidence rating), which outputs the text results.

AWS Transcribe

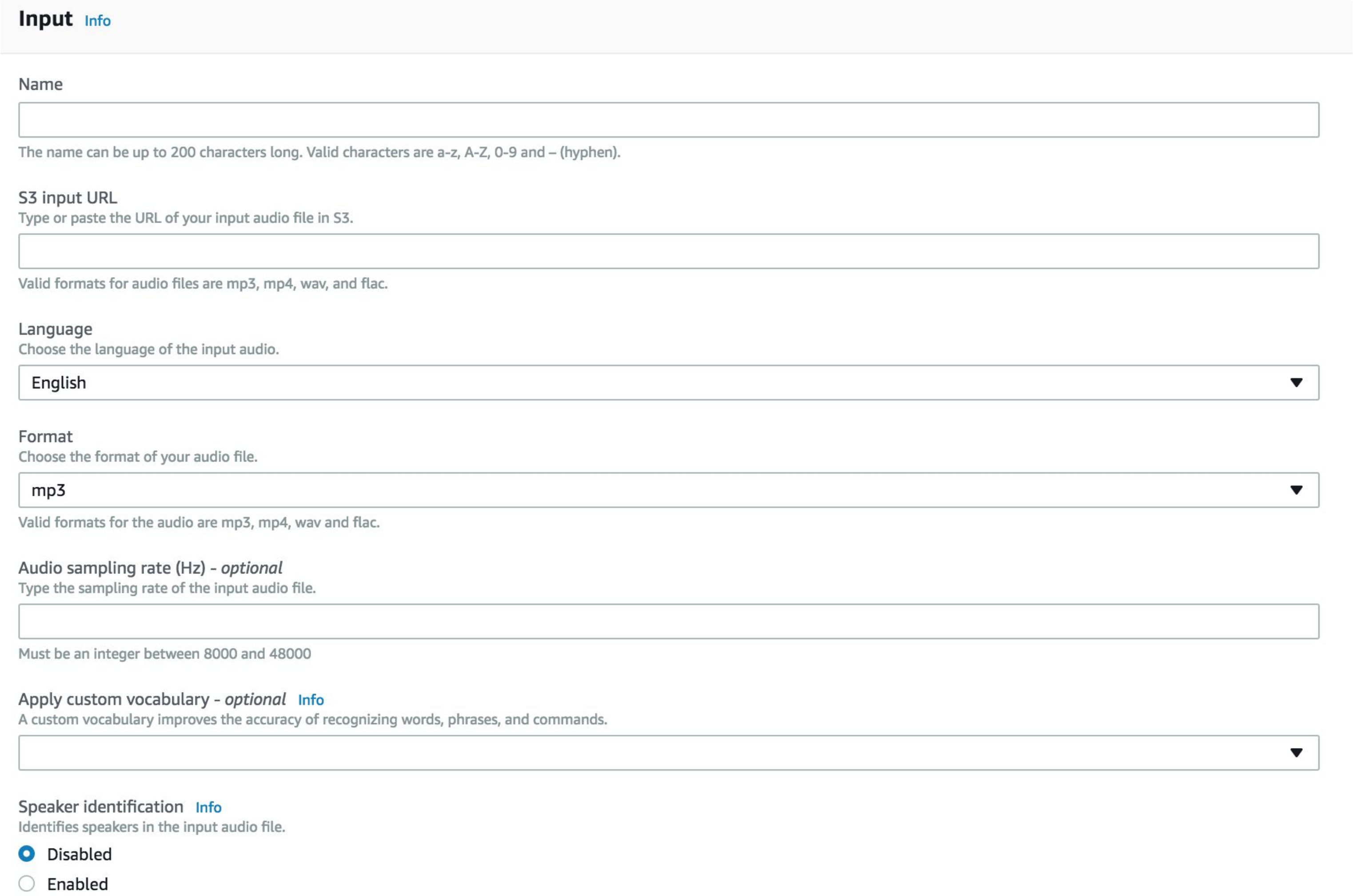
Currently available in 4 AWS AZs (not in Australia yet).

You can use the console to directly talk to the service, use the Amazon CLI, or utilise the APIs to trigger a service.

You can add custom vocabularies via CSV (important feature) and use speaker identification.

Results are stored for 90 days.

Cost: 60 mins / month for 12 months free.
Then $0.0004 second (approx. $AUD 1.80 hour)



Input Info

Name

The name can be up to 200 characters long. Valid characters are a-z, A-Z, 0-9 and – (hyphen).

S3 input URL

Type or paste the URL of your input audio file in S3.

Valid formats for audio files are mp3, mp4, wav, and flac.

Language

Choose the language of the input audio.

English ▼

Format

Choose the format of your audio file.

mp3 ▼

Valid formats for the audio are mp3, mp4, wav and flac.

Audio sampling rate (Hz) - *optional*

Type the sampling rate of the input audio file.

Must be an integer between 8000 and 48000

Apply custom vocabulary - *optional* Info

A custom vocabulary improves the accuracy of recognizing words, phrases, and commands.

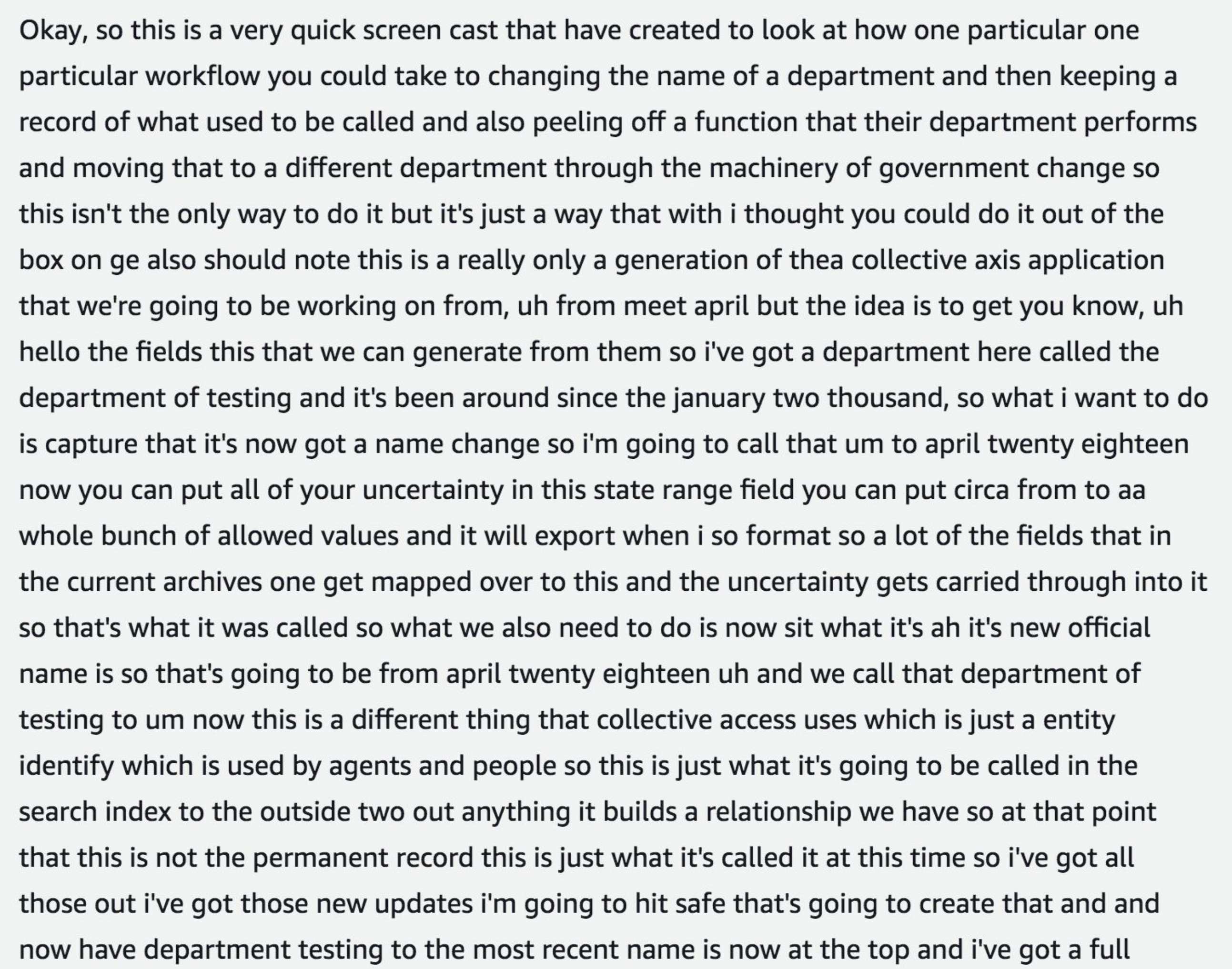
▼

Speaker identification Info

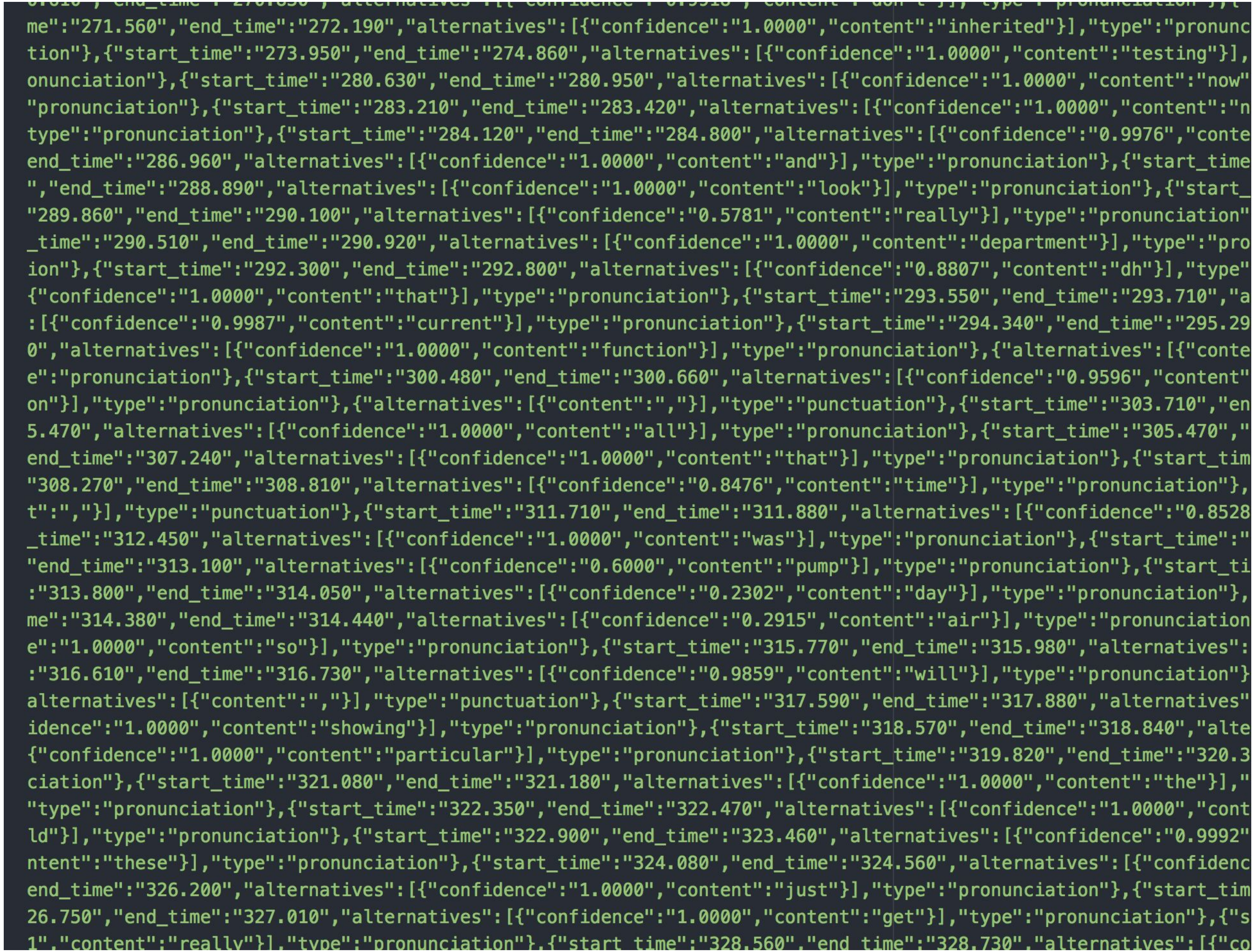
Identifies speakers in the input audio file.

● Disabled

○ Enabled



Okay, so this is a very quick screen cast that have created to look at how one particular one particular workflow you could take to changing the name of a department and then keeping a record of what used to be called and also peeling off a function that their department performs and moving that to a different department through the machinery of government change so this isn't the only way to do it but it's just a way that with i thought you could do it out of the box on ge also should note this is a really only a generation of thea collective axis application that we're going to be working on from, uh from meet april but the idea is to get you know, uh hello the fields this that we can generate from them so i've got a department here called the department of testing and it's been around since the january two thousand, so what i want to do is capture that it's now got a name change so i'm going to call that um to april twenty eighteen now you can put all of your uncertainty in this state range field you can put circa from to aa whole bunch of allowed values and it will export when i so format so a lot of the fields that in the current archives one get mapped over to this and the uncertainty gets carried through into it so that's what it was called so what we also need to do is now sit what it's ah it's new official name is so that's going to be from april twenty eighteen uh and we call that department of testing to um now this is a different thing that collective access uses which is just a entity identify which is used by agents and people so this is just what it's going to be called in the search index to the outside two out anything it builds a relationship we have so at that point that this is not the permanent record this is just what it's called it at this time so i've got all those out i've got those new updates i'm going to hit safe that's going to create that and and now have department testing to the most recent name is now at the top and i've got a full



```
0.010","end_time":"270.830","alternatives":[{"confidence":"0.9918","content":"don't"}],"type":"pronunciation"},{
me":"271.560","end_time":"272.190","alternatives":[{"confidence":"1.0000","content":"inherited"}],"type":"pronunc
tion"},{"start_time":"273.950","end_time":"274.860","alternatives":[{"confidence":"1.0000","content":"testing"}],
onunciation"},{"start_time":"280.630","end_time":"280.950","alternatives":[{"confidence":"1.0000","content":"now"
"pronunciation"},{"start_time":"283.210","end_time":"283.420","alternatives":[{"confidence":"1.0000","content":"n
type":"pronunciation"},{"start_time":"284.120","end_time":"284.800","alternatives":[{"confidence":"0.9976","conte
end_time":"286.960","alternatives":[{"confidence":"1.0000","content":"and"}],"type":"pronunciation"},{"start_time
","end_time":"288.890","alternatives":[{"confidence":"1.0000","content":"look"}],"type":"pronunciation"},{"start_
"289.860","end_time":"290.100","alternatives":[{"confidence":"0.5781","content":"really"}],"type":"pronunciation"
_time":"290.510","end_time":"290.920","alternatives":[{"confidence":"1.0000","content":"department"}],"type":"pro
ion"},{"start_time":"292.300","end_time":"292.800","alternatives":[{"confidence":"0.8807","content":"dh"}],"type"
{"confidence":"1.0000","content":"that"}],"type":"pronunciation"},{"start_time":"293.550","end_time":"293.710","a
:[{"confidence":"0.9987","content":"current"}],"type":"pronunciation"},{"start_time":"294.340","end_time":"295.29
0","alternatives":[{"confidence":"1.0000","content":"function"}],"type":"pronunciation"},{"alternatives":[{"conte
e":"pronunciation"},{"start_time":"300.480","end_time":"300.660","alternatives":[{"confidence":"0.9596","content"
on"}],"type":"pronunciation"},{"alternatives":[{"content":","}],"type":"punctuation"},{"start_time":"303.710","en
5.470","alternatives":[{"confidence":"1.0000","content":"all"}],"type":"pronunciation"},{"start_time":"305.470","
end_time":"307.240","alternatives":[{"confidence":"1.0000","content":"that"}],"type":"pronunciation"},{"start_tim
"308.270","end_time":"308.810","alternatives":[{"confidence":"0.8476","content":"time"}],"type":"pronunciation"},
t":","}],"type":"punctuation"},{"start_time":"311.710","end_time":"311.880","alternatives":[{"confidence":"0.8528
_time":"312.450","alternatives":[{"confidence":"1.0000","content":"was"}],"type":"pronunciation"},{"start_time":"
"end_time":"313.100","alternatives":[{"confidence":"0.6000","content":"pump"}],"type":"pronunciation"},{"start_ti
:"313.800","end_time":"314.050","alternatives":[{"confidence":"0.2302","content":"day"}],"type":"pronunciation"},
me":"314.380","end_time":"314.440","alternatives":[{"confidence":"0.2915","content":"air"}],"type":"pronunciation
e":"1.0000","content":"so"}],"type":"pronunciation"},{"start_time":"315.770","end_time":"315.980","alternatives":
:"316.610","end_time":"316.730","alternatives":[{"confidence":"0.9859","content":"will"}],"type":"pronunciation"}
alternatives":[{"content":","}],"type":"punctuation"},{"start_time":"317.590","end_time":"317.880","alternatives"
idence":"1.0000","content":"showing"}],"type":"pronunciation"},{"start_time":"318.570","end_time":"318.840","alte
{"confidence":"1.0000","content":"particular"}],"type":"pronunciation"},{"start_time":"319.820","end_time":"320.3
ciation"},{"start_time":"321.080","end_time":"321.180","alternatives":[{"confidence":"1.0000","content":"the"}],"
"type":"pronunciation"},{"start_time":"322.350","end_time":"322.470","alternatives":[{"confidence":"1.0000","cont
ld"}],"type":"pronunciation"},{"start_time":"322.900","end_time":"323.460","alternatives":[{"confidence":"0.9992"
ntent":"these"}],"type":"pronunciation"},{"start_time":"324.080","end_time":"324.560","alternatives":[{"confidenc
end_time":"326.200","alternatives":[{"confidence":"1.0000","content":"just"}],"type":"pronunciation"},{"start_tim
26.750","end_time":"327.010","alternatives":[{"confidence":"1.0000","content":"get"}],"type":"pronunciation"},{"s
1","content":"really"}],"type":"pronunciation"},{"start_time":"328.560","end_time":"328.730","alternatives":[{"co
```

AWS Transcribe results

Not too bad. There were lots of mistakes, but many could have been fixed with the vocabularies.

It gives a confidence rating on a word-by-word basis, which is good. But some that were > 99% confidence were wrong.

Interestingly, it actually improved as the audio went on - particularly in identifying sentence clauses.

But in all, it did a pretty good job and would have saved a lot of time.

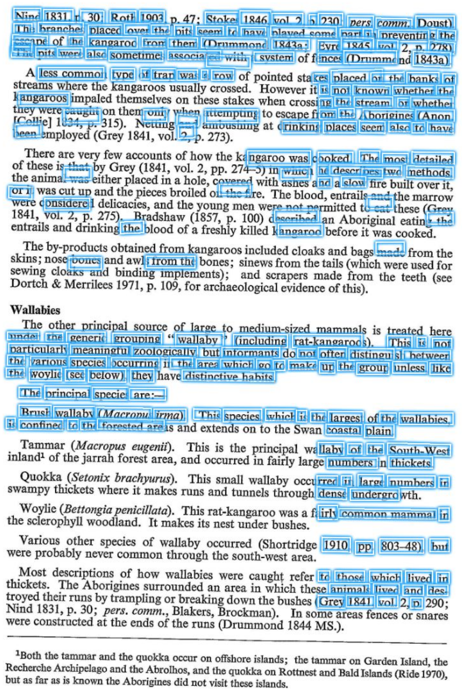
AWS Transcribe results

Took around 10 mins to process.

Couldn’t find how to produce a SRT compliant version.

But in all, it did a pretty good job and would have saved a lot of time.

Rekognition results



Nind 1831, p. 30; Roth 1903, p. 47; Stokes 1846, vol. 2, p. 230; *pers. comm.*, Doust). The branches placed over the pits seem to have played some part in preventing the escape of the kangaroo from them (Drummond 1843a; Eyre 1845, vol. 2, p. 278). The pits were also sometimes associated with a system of fences (Drummond 1843a).

A less common type of trap was a row of pointed stakes placed on the banks of streams where the kangaroos usually crossed. However it is not known whether the kangaroos impaled themselves on these stakes when crossing the stream, or whether they were caught on them only when attempting to escape from the Aborigines (Anon. [Collie] 1834, p. 315). Netting and ambushing at drinking places seem also to have been employed (Grey 1841, vol. 2, p. 273).

There are very few accounts of how the kangaroo was cooked. The most detailed of these is that by Grey (1841, vol. 2, pp. 274–5) in which he describes two methods, the animal was either placed in a hole, covered with ashes and a slow fire built over it, or it was cut up and the pieces broiled on the fire. The blood, entrails and the marrow were considered delicacies, and the young men were not permitted to eat these (Grey 1841, vol. 2, p. 275). Bradshaw (1857, p. 100) described an Aboriginal eating the entrails and drinking the blood of a freshly killed kangaroo before it was cooked.

The by-products obtained from kangaroos included cloaks and bags made from the skins; nose bones and awls from the bones; sinews from the tails (which were used for sewing cloaks and binding implements); and scrapers made from the teeth (see Dortch & Merrilees 1971, p. 109, for archaeological evidence of this).

Wallabies

The other principal source of large to medium-sized mammals is treated here under the generic grouping "wallaby" (including rat-kangaroos). This is not particularly meaningful zoologically, but informants do not often distinguish between the various species occurring in the area which go to make up the group unless, like the woylie (see below), they have distinctive habits.

The principal species are:—

Brush wallaby (*Macropus irma*). This species, which is the largest of the wallabies, is confined to the forested areas and extends on to the Swan coastal plain.

Tammar (*Macropus eugenii*). This is the principal wallaby of the South-West inland[1] of the jarrah forest area, and occurred in fairly large numbers in thickets.

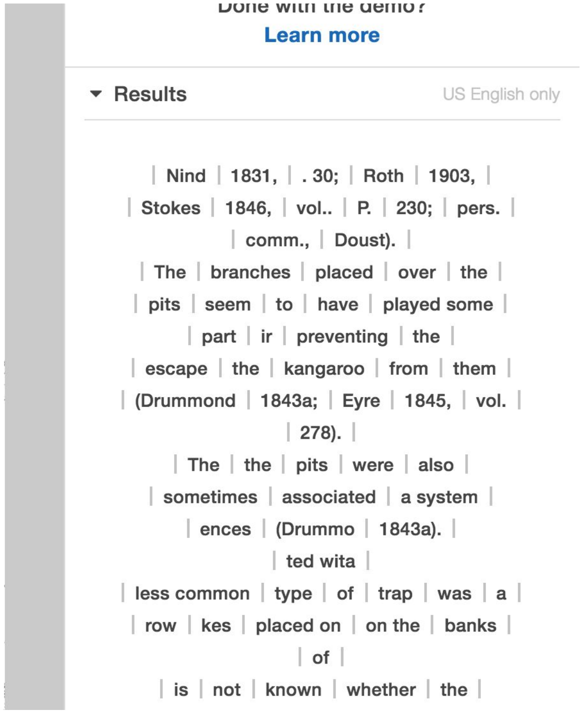
Quokka (*Setonix brachyurus*). This small wallaby occurred in large numbers in swampy thickets where it makes runs and tunnels through dense undergrowth.

Woylie (*Bettongia penicillata*). This rat-kangaroo was a fairly common mammal in the sclerophyll woodland. It makes its nest under bushes.

Various other species of wallaby occurred (Shortridge 1910, pp. 803–48), but were probably never common through the south-west area.

Most descriptions of how wallabies were caught refer to those which lived in thickets. The Aborigines surrounded an area in which these animals lived and destroyed their runs by trampling or breaking down the bushes (Grey 1841, vol. 2, p. 290; Nind 1831, p. 30; *pers. comm.*, Blakers, Brockman). In some areas fences or snares were constructed at the ends of the runs (Drummond 1844 MS.).

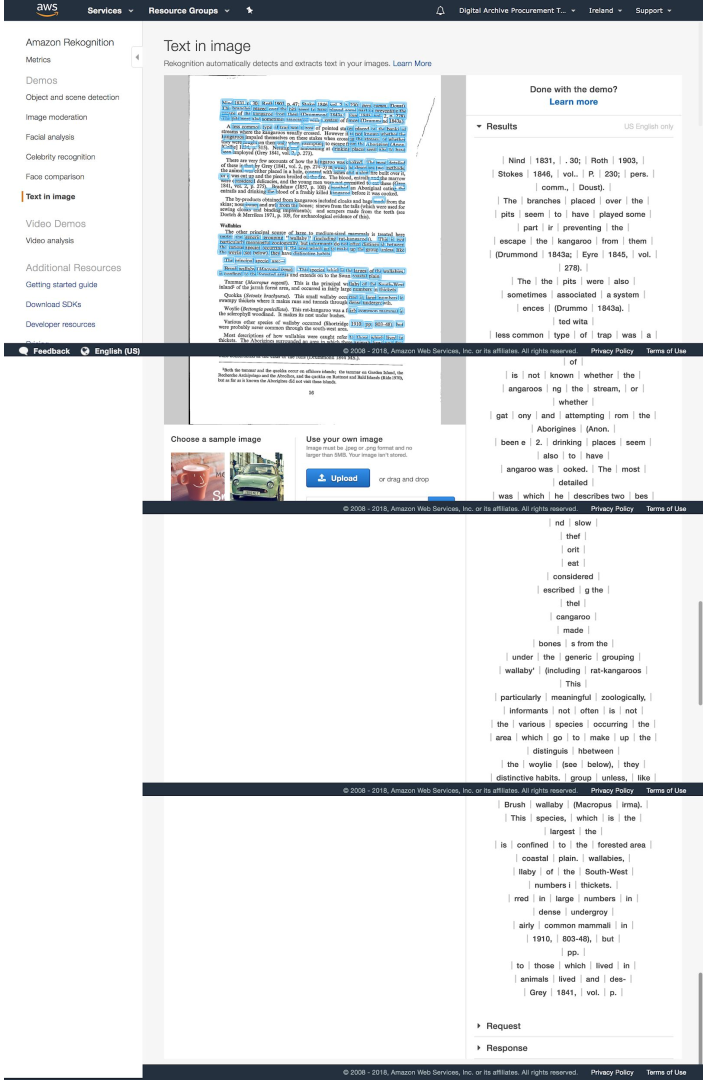
[1] Both the tammar and the quokka occur on offshore islands; the tammar on Garden Island, the Recherche Archipelago and the Abrolhos, and the quokka on Rottnest and Bald Islands (Ride 1970), but as far as is known the Aborigines did not visit these islands.



Results

US English only

```
| Nind | 1831, | . 30; | Roth | 1903, |
| Stokes | 1846, | vol.. | P. | 230; | pers. |
| comm., | Doust). |
| The | branches | placed | over | the |
| pits | seem | to | have | played some |
| part | ir | preventing | the |
| escape | the | kangaroo | from | them |
| (Drummond | 1843a; | Eyre | 1845, | vol. |
| 278). |
| The | the | pits | were | also |
| sometimes | associated | a system |
| ences | (Drummo | 1843a). |
| ted wita |
| less common | type | of | trap | was | a |
| row | kes | placed on | on the | banks |
| of |
| is | not | known | whether | the |
```



Text in image

Rekognition automatically detects and extracts text in your images. Learn More

Results

US English only

```
| Nind | 1831, | . 30; | Roth | 1903, |
| Stokes | 1846, | vol.. | P. | 230; | pers. |
| comm., | Doust). |
| The | branches | placed | over | the |
| pits | seem | to | have | played some |
| part | ir | preventing | the |
| escape | the | kangaroo | from | them |
| (Drummond | 1843a; | Eyre | 1845, | vol. |
| 278). |
| The | the | pits | were | also |
| sometimes | associated | a system |
| ences | (Drummo | 1843a). |
| ted wita |
| less common | type | of | trap | was | a |
| of |
| is | not | known | whether | the |
```

```
| angaroos | ng | the | stream, | or |
| whether |
| gat | ony | and | attempting | rom | the |
| Aborigines | (Anon. |
| been e | 2. | drinking | places | seem |
| also | to | have |
| angaroo was | ooked. | The | most |
| detailed |
| was | which | he | describes two | bes |
| nd | slow |
| thef |
| orit |
| eat |
| considered |
| escribed | g the |
| thel |
| cangaroo |
| made |
| bones | s from the |
| under | the | generic | grouping |
| wallaby' | (including | rat-kangaroos |
| This |
| particularly | meaningful | zoologically, |
| informants | not | often | is | not |
| the | various | species | occurring | the |
| area | which | go | to | make | up | the |
| distinguis | hbetween |
| the | woylie | (see | below), | they |
| distinctive habits. | group | unless, | like |
| Brush | wallaby | (Macropus | irma). |
| This | species, | which | is | the |
| largest | the |
| is | confined | to | the | forested area |
| coastal | plain. | wallabies, |
| llaby | of | the | South-West |
| numbers i | thickets. |
| rred | in | large | numbers | in |
| dense | undergroy |
| airly | common mammali | in |
| 1910, | 803-48), | but |
| pp. |
| to | those | which | lived | in |
| animals | lived | and | des- |
| Grey | 1841, | vol. | p. |
```

Choose a sample image

Use your own image

Image must be .jpeg or .png format and no larger than 5MB. Your image isn't stored.

Discussion and conclusion

There is value here. You probably can’t rely on it as an end-to-end solution without checking, but it’s pretty good.

Use the custom vocabularies, and make it part of the workflow.

Also, don’t use Rekognition for things it’s not meant to do. It’s not a replacement OCR service.

Discussion and conclusion

Would it pass ethics for qualitative research? Probably not.

Will it replace transcription outsourcing? Not yet.

Can it reduce the cost of making accessible resources? Yes.

Is Google’s service superior to Amazon’s? At the minute it is.

REPEAT MY DISCLAIMER!!!!

Before I started writing this talk, I had never used either service.

I skim read the documentation after putting my son to bed on Sunday night.

My test is also statistically invalid - three tests in total.

I am not an expert in either service, I am merely road testing, and reporting on the result that I experienced.

Any questions?

Morgan Strong
Technical PM
Gaia Resources

@mostrorec
mostro20@gmail.com