CODIFY

RGUHack 2018

Speech recognition technology continues to improve – driven in part by natural language interfaces to computers such as Alexa, Siri, Google and Cortana.

Could this technology be used to automatically transcribe meetings? This would be a great way to capture what was said and agreed at business meetings without the time-consuming process of writing up meeting minutes. It would be a real-time saver for businesses of all kinds.

This is our challenge for the 2018 RGUHack. There are two parts:

1 Transcribe a pre-recorded meeting

We've provided an excerpt from the recent Zuckerberg trial… erm I mean "hearing". It is just over 2 minutes long. The challenge is to transcribe this into something close to the following output that was provided by the Washington Post[1]:

> WICKER: Is it true that — as was recently publicized, that Facebook collects the call and text histories of its users that use Android phones?
>
> ZUCKERBERG: Senator, we have an app called Messenger for sending messages to your Facebook friends. And that app offers people an option to sync their — their text messages into the messaging app, and to make it so that — so basically so you can have one app where it has both your texts and — and your Facebook messages in one place.
>
> We also allow people the option of ...
>
> WICKER: You can opt in or out of that?
>
> ZUCKERBERG: Yes. It is opt-in.
>
> WICKER: It is easy to opt out?
>
> ZUCKERBERG: It is opt-in. You — you have to affirmatively say that you want to sync that information before we get access to it.
>
> WICKER: Unless you — unless you opt in, you don't collect that call and text history?
>
> ZUCKERBERG: That is correct.
>
> WICKER: And is that true for — is this practice done at all with minors, or do you make an exception there for persons aged 13 to 17?
>
> ZUCKERBERG: I do not know. We can follow up with that.
>
> WICKER: Okay, do that — let's do that.
>
> One other thing: There have been reports that Facebook can track a user's Internet browsing activity, even after that user has logged off of the Facebook platform. Can you confirm whether or not this is true?

[1] https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/

ZUCKERBERG: Senator — I — I want to make sure I get this accurate, so it would probably be better to have my team follow up afterwards.

WICKER: You don't know?

ZUCKERBERG: I know that the — people use cookies on the Internet, and that you can probably correlate activity between — between sessions.

We do that for a number of reasons, including security, and including measuring ads to make sure that the ad experiences are the most effective, which, of course, people can opt out of. But I want to make sure that I'm precise in my answer, so let me ...

WICKER: When — well, when you get ...

ZUCKERBERG: ... follow up with you on that.

WICKER: ... when you get back to me, sir, would you also let us know how Facebook's — discloses to its users that engaging in this type of tracking gives us that result?

ZUCKERBERG: Yes.

WICKER: And thank you very much.

GRASSLEY: Thank you, Senator Wicker.

2 Transcribe a meeting in real-time

The natural next step would be to transcribe the meeting as it happens in near real-time. Whenever there is a pause in speaking, the audio fragment could be transcribed and added to the output.

3 Implementation thoughts

The major technology players have cloud APIs for speech recognition. For example:

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

https://azure.microsoft.com/en-gb/services/cognitive-services/directory/speech/

https://cloud.google.com/speech-to-text/

Operating systems also have speech APIs that might be more appropriate for the near real-time scenario. For example on Microsoft Windows 10:

Speech Recognition API: https://msdn.microsoft.com/en-us/library/jj127860.aspx