# Hey ArchivistaBox, who said what, when and where?

***Egg, 25 May 2021:*** *With the integration of speech recognition in the ArchivistaBox 2021/V, the spoken text can now be extracted from any sound and video files. The following blog is about showing why speech recognition makes a lot of sense for the ArchivistaBox and how easily it works.*


## Open Source Speech Recognition Vosk and Kaldi

Until now, captured audio and video files could only be automatically tagged in a rudimentary way. Thanks to the integration of the Vosk speech recognition system, it is now possible to convert spoken passages from the multimedia files into text and to prepare this for the ArchivistaBox full text search. The languages English, German, French, Italian, Spanish, Portuguese and Dutch are currently supported.

Thanks to the technologies of the open source language recognizers **Vosk** and **Kaldi,** further languages can be added at any time. It would also be possible to create your own (new) language files and integrate them into the ArchivistaBox. In most cases this will not be necessary, as there are already approx. 10 GBytes (unpacked) of vocabulary available for the above languages in the basic scope.

At this point it may be added that the integration of speech recognition is done via Python and the corresponding Vosk module. Vosk in turn is based on Kaldi. It should be noted that Kaldi forms the actual basis, while Vosk simplifies the job considerably. Without Vosk, Kaldi would first have to be trained with speech sequences. Thanks to Vosk, Kaldi is ultimately called up in such a way that the text can be automatically extracted from the video and audio files.

## Two examples with recognized text

In order to illustrate the quality of speech recognition, an excerpt from **SRF's Tagesschau of 23 May 2021 (inkl. translation with deepl.com)** is published here (it is about the volcanic eruption in the Congo on the same day): ***"Thousands are still leaving the area around the city of Goma in the night towards the border with Rwanda – the neighboring country has opened the borders for the volcanic refugees and is providing emergency accommodation".***

The quality of recognition may be seen (absolutely no training effort was made). Normal words are almost always recognized without error, but upper and lower case letters are not, and errors can occur with place names (same program, cable car accident in Italy): ***"First to Italy in the north of the country, a cable car cabin has crashed killing thirteen people the accident occurred near the Swiss border on the journey from the vacation resort of strehla on the shores of lago maggiore to the monte matrone…"*** Instead of 'Strehla' it should be 'Stresa' and Monte Mattarone was renamed Matrone. On the other hand, Lago Maggiore could be correctly identified.

## Installation of Vosk and Kaldi

Good speech recognition requires a relatively large amount of space for the speech files. For this reason, Vosk or Kaldi is not located directly on the ArchivistaBox ISO file. However, the corresponding download link for the ‘vosk.os’ file is readily communicated to customers. This file must be copied to ‘/home/data’. The ArchivistaBox 2021/V must then be restarted. This activates the speech recognition.

When adding content, the language recognition is carried out directly after the text recognition. The desired language must be specified in the scan definition. In principle, speech recognition works on all ArchivistaBox systems. However, it should be noted that the Dolder, Rigi and Säntis models will work somewhat slower than the other models.

Speech recognition takes place within the framework of text recognition (OCR). If this is processed in parallel, speech recognition is also performed with several instances. The recognition speed (per processor) is approximately 5 to 10 minutes for one hour of audio. With the ArchivistaBox MediaVM Everest, up to 2400 hours or 100 days of 24 hours of audio can be processed per day. The entire speech recognition process is performed directly on the local instance, which ensures 100% data confidentiality.

## Speech recognition is available to all customers

Although the new technology is not activated on delivery, it is available for all ArchivistaBox systems in the basic scope. The only prerequisite is that at least the current version 2021/V is used or that the file ‘vosk.os’ is present in the folder ‘/home/data’.

Vosk and Kaldi can also be installed at customer sites on request and adapted to customer-specific requirements. The following applies as always in the ArchivistaBox cosmos. Extensions that are created as part of an order are (unless the reverse is explicitly agreed) later available to all customers. With this in mind, have fun with the new speech recognition on the ArchivistaBox.