Hey ArchivistaBox, wer hat was, wann und wo gesagt?

Egg, 25. Mai 2021: *Mit der Integration der Spracherkennung in die ArchivistaBox 2021/V kann neu von beliebigen Ton- und Video-Dateien der gesprochene Text extrahiert werden. In nachfolgendem Blog geht es darum aufzuzeigen, warum die Spracherkennung für die ArchivistaBox viel Sinn ergibt und wie einfach diese funktioniert.*

Open Source Spracherkennungen Vosk und Kaldi

Bisher konnten erfasste Ton- und Video-Dateien nur rudimentär automatisiert beschlagwortet werden. Dank der Integration der Spracherkennung Vosk ist es neu möglich, gesprochene Passagen aus den multimedialen Dateien in Text umzuwandeln und diesen für die Folltextrecherche der ArchivistaBox aufzubereiten. Dabei werden aktuell die Sprachen Englisch, Deutsch, Französisch, Italienisch, Spanisch, Portugiesisch und Holländisch unterstützt.

Dank der Technologien der Open Source Spracherkennungen **Vosk** und **Kaldi** können jederzeit weitere Sprachen hinzugefügt werden. Ebenfalls wäre es möglich, eigene (neue) Sprachdateien zu erstellen und in die ArchivistaBox zu integrieren. In den meisten Fällen wird dies nicht notwendig sein, stehen im Grundumfang bereits ca. 10 GByte (entpackt) für die obigen Sprachen an Vokabular zur Verfügung.

An dieser Stelle darf angefügt werden, dass die Integration der Spracherkennung über Python und das entsprechende Vosk-Modul erfolgt. **Vosk** wiederum basiert auf **Kaldi.** Dabei gilt es anzumerken, dass Kaldi die eigentliche Grundlage bildet, Vosk hingegen den Job deutlich vereinfacht. Ohne Vosk müsste Kaldi erst mit Sprachsequenzen trainiert werden. Dank Vosk wird Kaldi letztlich so aufgerufen, dass aus den Video- und

Audio-Dateien der Text automatisiert extrahiert werden kann.

Zwei Beispiele mit erkanntem Text

Um die Qualität der Spracherkennung zu veranschaulichen, sei an dieser Stelle eine Auszug aus der Tagesschau von **SRF vom 23. Mail 2021** publiziert (es geht um den Vulkanausbruch im Kongo vom gleichen Tag): «***Noch in der nacht verlassen tausende das gebiet um die großstadt goma in richtung grenze nach ruanda das nachbarland hat die grenzen für die vulkan flüchtlinge geöffnet und stellt notunterkünfte zur verfügung.»***

Die Qualität der Erkennung darf sich (es wurde absolut kein Trainingsaufwand betrieben) sehen lassen. Normale Wörter werden fast immer fehlerfrei erkannt, Gross- und Kleinschreibung dagegen nicht und bei Ortsbezeichnungen kann es zu Fehlern kommen (Gleiche Sendung, Seilbahn-Unglück in Italien): ***«Zuerst nach italien im norden des landes ist eine seilbahn kabine abgestürzt dabei sind dreizehn menschen ums leben gekommen das unglück ereignete sich nahe der schweizer grenze auf der fahrt vom ferienort strehla am ufer des lago maggiore auf den monte matrone…»*** Anstelle von ‹Strehla› müsste ‹Stresa› stehen und der Monte Mattarone wurde in Matrone umbenannt. Dagegen konnte der Lago Maggiore korrekt erkannt werden.

Installation von Vosk und Kaldi

Eine gute Spracherkennung benötigt relativ viel Platz für die Sprachdateien. Daher befindet sich Vosk bzw. Kaldi nicht direkt auf der ISO-Datei der ArchivistaBox. Der entsprechende Download-Link der Datei ‹vosk.os› wird Kunden aber gerne kommuniziert. Diese Datei ist nach ‹/home/data› zu kopieren. Danach ist die ArchivistaBox 2021/V neu zu starten. Damit wird die Spracherkennung aktiviert.

Beim Hinzufügen von Inhalten wird die Spracherkennung direkt nach der Texterkennung durchgeführt. Die gewünschte Sprache ist dabei über die Scan-Definition festzulegen. Grundsätzlich arbeitet die Spracherkennung auf sämtlichen ArchivistaBox-Systemen. Allerdings gilt es zu bedenken, dass die Modelle Dolder, Rigi und Säntis etwas langsamer arbeiten werden als die übrigen Modelle.

Die Spracherkennung erfolgt im Rahmen der Texterkennung (OCR). Wird diese parallel abgearbeitet, erfolgt auch die Spracherkennung mit mehreren Instanzen. Die Erkennungsgeschwindigkeit (pro Prozessor) beträgt ca. 5 bis 10 Minuten für eine Stunde Ton. Mit der ArchivistaBox MediaVM Everest können dabei pro Tag bis zu 2400 Stunden bzw. 100 Tage à 24 Stunden Ton verarbeitet werden. Der gesammte Prozess der Spracherkennung erfolgt dabei direkt auf der lokalen Instanz, womit Vertraulichkeit über die Daten zu 100 Prozent gewährleistet ist.

Spracherkennung steht allen Kunden zur Verfügung

Die neue Technologie ist zwar bei der Auslieferung nicht aktiviert, sie steht aber für alle ArchivistaBox-Systeme im Grundumfang zur Verfügung. Voraussetzung ist einzig, dass minimal die aktuelle Version 2021/V verwendet wird bzw. dass die Datei ‹vosk.os› im Ordner ‹/home/data› vorhanden ist.

Vosk und Kaldi können auf Wunsch auch bei Kunden installiert und an kundenspezifische Bedürfnisse angepasst werden. Dabei gilt wie immer im ArchivistaBox-Kosmos. Erweiterungen, die im Rahmen eines Auftrages entstehen, stehen (sofern nicht explizit das Gegenteil vereinbart wird) später allen Kunden zur Verfügung. In diesem Sinne viel Spass mit der neuen Spracherkennung auf der ArchivistaBox.