Richtlinie zum „Software-Sprint“

Transcribee - Schaffung einer Web-Plattform für gemeinsame Audio- und Video-Transkription

Schlussbericht

Zuwendungsempfänger:

Karl Engelhardt, Jaro Habiger, Robin Heinemann und Philipp Mandler GbR

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS23S14 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Durch Transkripte können in sprachlastigen Formaten wie Podcasts und Videos Barrieren abgebaut werden. Hierzu ist es wünschenswert, hochqualitative Transkriptionen von Audio- und Video-Aufnahmen zu erstellen. Doch gerade bei längeren Aufnahmen ist die Transkription ein immenser Aufwand, der entweder viel Zeit kostet oder schlicht nicht geleistet werden kann. Die Transkription als Dienstleistung zuzukaufen ist für viele, gerade ehrenamtliche, Produktionen zu teuer, weshalb auf ihre Erstellung zulasten von Barrierefreiheit und Durchsuchbarkeit verzichtet wird. Auch qualitative Forschung benötigt oft als Schritt zur strukturierten Analyse gute Interview-Transkripte. Hier stehen die Forscher*innen vor dem gleichen Problem: Diese zu erstellen bedeutet entweder großen Aufwand oder hohe Kosten für professionelle Transkriptionsdienstleister.

Bei der Transkription kann mithilfe von moderner Computersoftware viel Arbeit gespart werden. Um die Erstellung von Transkripten zu erleichtern haben wir eine Webplattform entwickelt, die es ermöglicht Audio- und Videoaufnahmen großteils automatisch in Textform übersetzen zu lassen. Die automatische Transkription erfolgt dabei durch ein State-of-the-art AI-Modell. Die Qualität der Transkripte dieser Modelle ist in jüngster Zeit stark gestiegen, jedoch treten trotzdem vereinzelt Fehler auf. Die Korrektur des automatischen Transkriptes ist daher ein wichtiger Bestandteil der Plattform und erfolgt kollaborativ (ähnlich zu Google Docs) und kann durch mehrere Personen gleichzeitig an separaten Arbeitsplätzen erfolgen.

Die Meilensteine, die wir für die Entwicklung festgelegt haben, waren:

1. Aufbau der Grundfunktionen (Grundlage WebApp/Backend inkl. Netzwerkkommunikation, Login, Anlage von Dokumenten, Worker-Architektur, automatische Transkription)
2. Export von Transkripten, kollaboratives Arbeiten am Dokument, automatische zeitliche Zuordnung von Text zu Ton (Alignment)

3. Funktion zur manuellen Korrektur des Alignments

Beitrag des Projektes zu den Zielen der Förderinitative „Software-Sprint“

Transkripte von Tonaufnahmen oder Videos wie z. B. Vorträgen, Reden oder Pressekonferenzen sind für die Barrierefreiheit dieser essentiell. Die textuelle Repräsentation multimedialer Formate ermöglicht es außerdem die Inhalte dieser multimedialen Formate zu durchsuchen oder überhaupt erst aufzufinden (z. B. Dank der Indizierung des Textes durch eine Suchmaschine). Transcribee vereinfacht die Erstellung von Transkripten und ermöglicht so in diesen beiden Feldern (Barrierefreiheit und Auffindbarkeit) gezielte Verbesserungen.

Ausführliche Darstellung der Ergebnisse

Im Rahmen der Förderung konnte eine Webplattform entwickelt werden, mit der Mediendateien automatisch transkribiert werden können. Zudem werden automatisch Zeitmarken für den transkribierten Text berechnet. So entstandene Transkripte können in einem zweiten Schritt manuell und kollaborativ korrigiert werden, wofür ein interaktiver Editor für Transkripte entwickelt wurde, in dem das Original-Audio abgespielt und der zugehörige Text gleichzeitig hervorgehoben wird. Funktionen wie Zurückspulen, unterschiedliche Abspielgeschwindigkeiten oder an eine andere Stelle des Audios anhand des Textes springen wurden implementiert. Stellen des Transkriptes mit niedriger Erkennungssicherheit werden hervorgehoben um die Korrektur einfach und effizient zu gestalten.

Zusätzlich zur automatischen Transkription wird automatisch erkannt, an welchen Stellen im Dokument die gleiche Person spricht. Auch diese Personenzuordnungen können korrigiert und mit Namen versehen werden. Die rechenintensiven Machine-Learning (ML) Schritte der Transkription können auf mehrere Computer verteilt werden um einen zuverlässigen und performanten Betrieb der Webplattform für mehrere Benutzer*innen zu gewährleisten.

Transkriptionsprojekte können auf der Webplattform angelegt, gelöscht und mit anderen Nutzer*innen geteilt werden (sowohl read-only als auch read-write). Das Transkript kann in verschiedenen Formaten exportiert werden, auch als Archiv, die auf anderen Transcribee-Instanzen importiert werden können.

Nicht implementiert werden konnte das nachträgliche Forced Re-Alignment von Texten sowie die manuelle Korrektur der generierten Zeitmarken.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Das Projekt kann von allen genutzt werden, die Audioaufnahmen transkribieren wollen. Dabei haben wir uns im Rahmen der Förderung vor allem auf zwei Gruppen konzentriert: Produzent*innen von Podcasts und Videos sowie Forscher*innen. Als das Projekt in einem nutzbaren Zustand war, haben wir ersten Nutzer*innen aus diesen Gruppen Zugriff auf eine von uns betriebene Transcribee-Instanz gegeben, um so Rückmeldungen zu bekommen und Probleme sowie Chancen in der weitern Umsetzung zu identifizieren.

Durch die Open-Source-Stellung können andere Entwickler*innen auf unserem Code aufsetzen und weitergehende Lösungen entwickeln. Bereits während der Förderzeit konnten wir Gespräche mit Entwickler*innen führen, die dies planen.

Zusammen mit diesen wollen wir nach Möglichkeiten suchen, die Entwicklung fortzuführen und sie in ihren Vorhaben zu unterstützen. Zudem planen wir Transcribee als SaaS-Lösung anzubieten.

Im Rahmen der Entwicklungszeit konnten wir neue Erfahrungen mit Technologien wie modernen Spracherkennungssystemen und kollaborativen Datentypen (CRDTs) sammeln. Diese wertvollen Erfahrungen werden in Zukunft anderen Open-Source-Projekten zu denen die Zuwendungsempfänger*innen beitragen zugutekommen.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Im Rahmen des Architekturdesigns der Plattform wurden verschiedene Ansätze diskutiert und verworfen. So wurde eine client-zentrierte sowie eine microservice-basierte Architektur sowie eine vollständig Ende-zu-Ende-verschlüsselung aufgrund des damit verbundenen Aufwands zunächst verworfen.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Der Source-Code ist auf https://github.com/bugbakery/transcribee einsehbar. Unter https://transcribee.net finden sich weitere Informationen zum Projekt. Der Social-Media-Account des Teams hinter Transcribee ist unter https://chaos.social/@bugbakery abrufbar.

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

In den Grundzügen konnte die Arbeits- und Kostenplanung von uns eigehalten werden. Jedoch stellten sich während des Förderzeitraums einige der von uns eingesetzten Bibliotheken (z. B. die CRDTs Y.js / Automerge, Slate.js) als weniger stabil bzw. performant als erwartet heraus. So mussten wir einige Zeit in Fehlersuche und Performance-Optimierung stecken, wodurch wir die nachträgliche Anpassung von Zeitmarken (sowohl automatisch, als auch manuell) nicht umsetzen konnten. Dies stellt jedoch keine signifikante Beeinträchtigung in der Benutzbarkeit der Webplattform dar, da es sich um sehr selten benötigte Funktionen handelt.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Aufgrund der Nutzung von Open-Source Bibliotheken haben wir bei der Entwicklung von Transcribee dazu beitragen, dass Probleme und Fehler in verschiedenen Libraries gefunden und behoben wurden. So haben wir z. B. Fehler in automerge, automerge-py, html5-audio-driver und whispercpp gefunden und behoben.