



netidee

PROJEKTE

Open Audio Search

Zwischenbericht | Call 15 | Projekt ID 5198

Lizenz: CC-BY-SA

Inhalt

1	Einleitung	3
2	Status der Arbeitspakete	3
2.1	Arbeitspaket 1 - Detailplanung und Formales am Projektstart	3
2.2	Arbeitspaket 2 - NLP	4
2.3	Arbeitspaket 3 - Trainingspipeline/Evaluationspipeline	4
2.4	Arbeitspaket 4 - Hardware	5
2.5	Arbeitspaket 5 - ASR-Pipeline an OAS-Core	6
2.6	Arbeitspaket 6 - Preprocessing & Analyse	6
2.7	Arbeitspaket 7 - Dokumentation und Formales während des Projekts und am Projektende ..	7
3	Umsetzung Förderauflagen	7
4	Zusammenfassung Planaktualisierung	7
5	Öffentlichkeitsarbeit/ Vernetzung	8
6	Eigene Projektwebsite	8

1 Einleitung

Auch wenn die zeitliche Abfolge der Arbeitspakete vom ursprünglichen Plan abweichen, liegen wir mit unserem Projekt Open Audio Search gut in der Zeit. Seit einiger Zeit schon ist das Projekt auf Github open sourced, die Hardware wurde installiert und in Betrieb genommen und auf <https://demo.openaudiosearch.org> ist eine erste Demo-Version unserer Software veröffentlicht worden.

2 Status der Arbeitspakete

2.1 Arbeitspaket 1 - Detailplanung und Formales am Projektstart

Kurzbeschreibung der Haupttätigkeiten

- Vertrag unterschrieben, Detailprojektplan (Arbeitsblatt Arbeitspakte) erstellt und abgenommen,
- detaillierte Liste Projektergebnisse mit Lizenz und Ort der öffentlichen Bereitstellung erstellt und abgenommen (Arbeitsblatt Projektergebnisse)
- Projekt-Website in Betrieb & erster Blogbeitrag erstellt
- erste Förderrate beantragt; ggf. nach Feedback ergänzt
- Förderrate1 genehmigt

Erkenntnisse zur Vorgangsweise

Keine besonderen Erkenntnisse.

Kurzbeschreibung der erreichten Ergebnisse

Siehe Projektplanung

Besondere Erfolge/ Probleme

Keine besonderen Erfolge und Probleme

Gab es große Abweichungen zum Plan? Warum?

Nein.

2.2 Arbeitspaket 2 - NLP

Kurzbeschreibung der Haupttätigkeiten

- NER Modul, das ASR Results nimmt und Entities erkennt.
- Punctuation Modul, das Transkripte nimmt und Interpunktionsfehler wiederherstellt.
- Erstellung eigener Knowledgebase (CBA Keywprds, Wikidata)
- Abgleich und Korrektur der Resultate mit Knowledgebase

Erkenntnisse zur Vorgangsweise

Kurzbeschreibung der erreichten Ergebnisse

Besondere Erfolge/ Probleme

Gab es große Abweichungen zum Plan? Warum?

Aufgrund mangelhafter zeitlicher Ressourcen ist der Fortschritt bei diesem Arbeitspaket noch nicht so weit, wie wir es ursprünglich geplant hätten. Da es aber keine technischen Abhängigkeiten zwischen den Arbeitspaketen gibt, stellt das kein schwerwiegendes Problem für den Gesamtfortschritt des Projektes dar.

2.3 Arbeitspaket 3 - Trainingspipeline/Evaluationspipeline

Kurzbeschreibung der Haupttätigkeiten

- Docker container, der die gesamte Trainingspipeline einschließt.
- Pfade für Daten bzw. Modelldateien etc. werden in dem Container gemounted.
- Grundsätzlich sollen existierende Kaldi-Skripte zum Trainieren (Bash, Python, ausführbare Dateien) wiederverwendet werden.
- Datapreparation
- Skripte zum automatischen Alignieren von Transkripten und Audiodaten
- Skripte zum Import von Daten und Konvertieren in ein Format, das für Training verwendet werden kann.
- Evaluationspipeline ist entwickelt und implementiert.

Erkenntnisse zur Vorgangsweise

- Die Trainingspipeline trainiert sogenannte "chain" Modelle, das sind akustische Modelle, die unter Verwendung von Kaldi mit lattice-free maximum mutual information (LF-MMI) Zielfunktion trainiert werden. Für diese Art von Modellen ist es erforderlich, sogenannte Alignments für die Trainingsdaten zu erzeugen. Um dies zu Bewerkstelligen, werden kleinere ASR-Modelle auf Basis der Daten trainiert, die im Anschluss zur Erzeugung der Alignments verwendet werden.

Kurzbeschreibung der erreichten Ergebnisse

- Eine Trainingspipeline auf Basis der derzeit modernsten Kaldi-Rezepte ("chain2") für das Trainieren von akustischen Modellen wurde entwickelt.

- Eine Pipeline zum Alignieren und Bereinigen von transkribierten Daten erstellt. Diese Pipeline ist in der Lage, Transkripte, die nur ungefähr mit den Audiodaten übereinstimmen, zu verarbeiten, um für das Trainieren von Modellen verwendbare Trainingsdaten zu erzeugen.
- Programme wurden erstellt, um größere Mengen neuer Daten für das Trainieren von ASR Modellen aus dem Internet herunterzuladen (zu "scrapen").
- Programme wurden erstellt, um existierende Korpora sowie große Mengen an neuen Daten für Training sowie Evaluierung vorzubereiten. Dies inkludiert Konvertieren der Dateien sowie Formatieren in das Kaldi-Datafolder Format.
- Evaluation von Wortfehlerrate ist derzeit als Teil der Trainingspipeline implementiert.

Besondere Erfolge/ Probleme

- Modelle basierend auf rel. geringen Datenmengen wurden trainiert, um die Funktionstüchtigkeit der Trainingspipeline zu Verifizieren.
- Das Trainieren von Modellen auf größeren Datenmengen, die eine höhere Performanz aufweisen, ist derzeit nicht möglich, da dafür vorgesehenen Grafikkarten noch nicht gekauft wurden.
- Derzeit ist die Trainingspipeline noch nicht in Docker container enkapsuliert.

Gab es große Abweichungen zum Plan? Warum?

- Bislang keine größeren Abweichungen

2.4 Arbeitspaket 4 - Hardware

Kurzbeschreibung der Haupttätigkeiten

- Hardware-Infrastruktur für Trainings- und Evaluationspipeline ist eingerichtet.

Erkenntnisse zur Vorgangsweise

Kurzbeschreibung der erreichten Ergebnisse

- Server wurde gekauft, zusammengebaut und eingerichtet
- Setup für Demo-Instanz wurde eingerichtet
- Demo-Instanz läuft unter <https://demo.openaudiosearch.org>
- Setup für Trainingspipeline ist vorbereitet

Besondere Erfolge/ Probleme

- Erfolge: Lauffähige Demo-Instanz (s.o.)
- Probleme: Bislang keine, abseits von üblichen kleineren Problemen im Setup-Prozess, die sich alle schnell lösen ließen

Gab es große Abweichungen zum Plan? Warum?

- Bislang wurden die Grafikkarten noch nicht gekauft aufgrund der aktuellen hohen Preissteigerungen

2.5 Arbeitspaket 5 - ASR-Pipeline an OAS-Core

Kurzbeschreibung der Haupttätigkeiten

- ASR-Pipeline ist entwickelt und in Core implementiert.

Erkenntnisse zur Vorgangsweise

- Der Umstieg auf Rust als Programmiersprache im Core hat sich als sehr sinnvoll erwiesen
- Die Arbeit mit dem Vosk-Toolkit als High-Level-Abstraction über Kaldi ist bislang zielführend

Kurzbeschreibung der erreichten Ergebnisse

- Inference-Engine via Vosk läuft im Core
- Anbindung an Modelle aus eigenem Training ist vorbereitet

Besondere Erfolge/ Probleme

- Implementierung in Rust war ein Erfolg
- Etablierung einer effizienten und leicht zu managenden Job-Queue zwischen Rust und Python erweist sich als Herausforderung, die noch nicht gänzlich zufriedenstellend gelöst ist
- Besondere Probleme gab es bislang keine

Gab es große Abweichungen zum Plan? Warum?

- Bislang keine größeren Abweichungen

2.6 Arbeitspaket 6 - Preprocessing & Analyse

Kurzbeschreibung der Haupttätigkeiten

- Software-Paket zur Segmentierung und Klassifikation von Audio-Dateien. Dieses Modul kann gesprochene Sprache effizient aus hochgeladenen Audiodateien extrahieren, hierfür wird die Audiodatei segmentiert und die Segmente werden als gesprochener Sprache, Musik oder als Hintergrundgeräusche klassifiziert.
- Software-Paket zur Diarisierung von gesprochener Sprache. Das Paket kann mit einer gewissen Fehlertoleranz gesprochene Sprache abstrakten Sprechern zuordnen.

Erkenntnisse zur Vorgangsweise

- Bislang in der Design- und Recherche phase

Kurzbeschreibung der erreichten Ergebnisse

- Planung zur Implementierung ist aufgestellt

Besondere Erfolge/ Probleme

- Bislang keine

Gab es große Abweichungen zum Plan? Warum?

- Bislang keine

2.7 Arbeitspaket 7 - Dokumentation und Formales während des Projekts und am Projektende

Kurzbeschreibung der Haupttätigkeiten

- Die geplanten Projektergebnisse (siehe Arbeitsblatt "Projektergebnisse") sind erstellt/ funktionsfähig und ausreichend dokumentiert;
- Projekt-Website wurde ein letztes Mal aktualisiert: Projektergebnisse sind unter Angabe der open source bzw. creative commons Lizenz der Öffentlichkeit zur Verfügung gestellt,
- Projektendbericht und Endabrechnung sind abgenommen; abschließender Projektblogeintrag; letzte Förderrate beantragt

Erkenntnisse zur Vorgangsweise

Kurzbeschreibung der erreichten Ergebnisse

Besondere Erfolge/ Probleme

Gab es große Abweichungen zum Plan? Warum?

3 Umsetzung Förderauflagen

Dieses Kapitel ist nur relevant, wenn in der Fördervereinbarung spezielle Förderauflagen festgelegt wurden. In diesem Fall soll in diesem Kapitel dargestellt werden, wie diese berücksichtigt werden.

Es wurden keine speziellen Förderauflagen vereinbart.

4 Zusammenfassung Planaktualisierung

Alle Anpassungen des Plan-Excels kurz zusammengefasst

Die Abgerechneten Stunden der Entwickler:innen wurden in die Stundenliste eingetragen, der Fortschritt der einzelnen Arbeitspakete aktualisiert. Kleinere Abweichungen bezüglich der geplanten finanziellen/zeitlichen Ressourcen ergeben sich einerseits durch die erhöhten Preise für Grafikkarten, andererseits durch den Umstand, dass die Aufwände für einzelne Arbeitspakete ursprünglich falsch eingeschätzt wurden.

5 Öffentlichkeitsarbeit/ Vernetzung

Beschreibung der bereits erfolgten Öffentlichkeitsarbeit oder Vernetzung, bzw. Beschreibung des Plans künftiger Aktivitäten

- Projektwebsite: <https://openaudiosearch.org>
- Demo: <https://demo.openaudiosearch.org>

Eine eigene Projektwebseite wurde eingerichtet, ebenso ist eine Demoversion online. Die Veröffentlichung dieser Demoversion wurde über die jeweiligen Social-Media Kanäle von cba und arso angekündigt. Ebenso wurden die Programmkoordinator:innen der freien Radios Österreichs per Mail-Aussendung informiert. Durch eine Teilnahme an einem NLP-Meetup konnte das Interesse von Entwickler:innen, die in diesem Bereich tätig sind, geweckt werden. Über den auf der Projektseite verlinkten Discord-Server wurden wir auch schon von Interessierten kontaktiert.

Ebenso konnte in persönlichen Gesprächen mit Menschen, die in den Bereichen Podcasting, Radio und Softwareentwicklung tätig sind, großes Interesse für Open Audio Search geweckt werden.

Mit Abschluss des Projekts und Fertigstellung der Software, werden die vorhandenen Kanäle von cba und arso genutzt, um die Öffentlichkeit zu informieren. Dabei werden gezielt Softwaredeveloper:innen, Radiomacher:innen, Podcaster:innen und Institutionen die Audio-Content produzieren, europaweit adressiert. Der Fokus wird hierbei jedoch auf den deutschsprachigen Raum gelegt.

6 Eigene Projektwebsite

Projektwebsite: <https://openaudiosearch.org>