Fachhochschule Südwestfalen
University of Applied Sciences

Bachelor- oder Projektarbeit

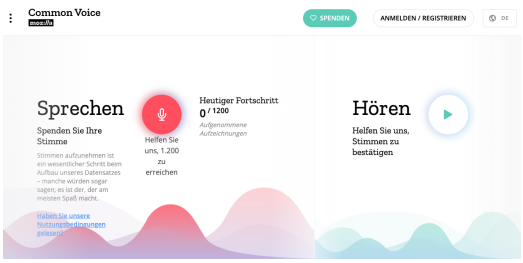
Deployment und Modifikation von Mozilla Common Voice zur Erfassung von Sprachdaten regionaler Dialekte

Iserlohn,
09.08.2024

Ziel der Arbeit

Im Bereich der automatischen Spracherkennung und Sprachsynthese ist die Verfügbarkeit geeigneter Trainingsdaten von entscheidender Bedeutung. Besonders bei der Entwicklung von KI-basierten Sprachmodellen stellt die Schaffung einer umfassenden Datenbasis eine große Herausforderung dar. Die quelloffene Webanwendung Common Voice wurde speziell für die Sammlung und Annotation von Sprachdaten entwickelt und bietet eine ideale Grundlage für derartige Vorhaben.

Ziel dieser Masterarbeit ist es, eine geeignete Datenbasis für das Training eines KI-basierten Sprachmodells für deutsche Dialekte (z.B. Sauerländer Platt) zu schaffen. Dazu wird die Webanwendung Common Voice lokal implementiert, angepasst und erweitert. Im Fokus steht die Modifikation des Quellcodes, um überflüssige Funktionen zu entfernen, geeignete Speicherlösungen zu verwenden und neue Sprachen hinzuzufügen.



Die Aufgaben der Arbeit umfassen:

- Lokales Deployment der quelloffenen Webanwendung Mozilla Common Voice.
- Bereitstellung der vorhandenen Funktionen (Textupload, Spracheingabe, Sprachausgabe, Bulk-Download) zur lokalen Nutzung.
- Reduktion der unterstützten Sprachen auf ein Minimum und Entfernung unnötiger Funktionen.
- Hinzufügung neuer Sprachen, insbesondere des Sauerländer Platt-Dialekts.
- Implementierung einer Benutzer- und Administratoranmeldung mit Profilen.
- Ermöglichung des Bulk-Uploads und der Segmentierung von Texten.

Quellen

[1] https://commonvoice.mozilla.org/de
[2] https://github.com/common-voice

Fachbereich Informatik und Naturwissenschaften

Prof. Dr.
Heiner Giefers

Cloud Computing

Telefon
02371 566-5252

E-Mail
giefers.heiner@fh-swf.de

Standort Iserlohn
Frauenstuhlweg 31
58644 Iserlohn

Fachhochschule
Südwestfalen
Sitz: Iserlohn

Hagen
Iserlohn
Lüdenscheid
Meschede
Soest

www.fh-swf.de