

BOT-13: Komponente zur Umwandlung von Sprache zu Text (STT)

Es soll ermöglicht werden, dass Benutzern neben Textnachrichten auch mittels Sprachnachrichten mit dem BeuthBot kommunizieren können. Dabei sollen die Sprachnachrichten mittels eines neuen Services in Text übersetzt werden und dann wie andere Textnachrichten verarbeitet werden. Hierzu sollen 3 bekannte STT-Frameworks (Kaldi, Mozilla Voice STT und Wav2Letter) getestet und verglichen werden. Basierend darauf soll eine Entscheidung getroffen werden, welches Framework schlussendlich in der Production-Environment verwendet werden soll. Das Framework wird dann in Form eines neuen Micro-Services in den BeuthBot integriert.

Zuständigkeit	Robert Halwaß
Initiale Schätzung	3 Tage
Programmiersprachen	* Python (Mozilla Voice STT) * C++ (Kaldi, WAV2Letter)
Frameworks/Libraries	* Kaldi * Mozilla Voice STT * WAV2Letter
Services	* Speech
Abhängigkeiten	* BOT-43: Erstellung eines Common-Frameworks für (Content-)Services
Anforderungen	* Die Übersetzung soll mittels neuronaler Netze geschehen * Nur Sprachnachrichten auf Deutsch sollen übersetzt werden * Das verwendete Framework muss OpenSource sein und Lokal auf dem BeuthBot-Server ausführbar sein * Es soll keine Model-Adaption durchgeführt werden
Tasks	* BOT-69: WAV2Letter testen * BOT-70: Mozilla Voice testen * BOT-73: Kaldi testen * BOT-71: Framework aussuchen * BOT-122: Micro-Service erstellen * BOT-123: Sprachnachricht in WAV umwandeln

Nutzungshinweis: Auf dieses vorliegende Schulungs- oder Beratungsdokument (ggf.) erlangt der Mandant vertragsgemäß ein nicht ausschließliches, dauerhaftes, unbeschränktes, unwiderrufliches und nicht übertragbares Nutzungsrecht. Eine hierüber hinausgehende, nicht zuvor durch *datenschutz-maximum* bewilligte Nutzung ist verboten und wird urheberrechtlich verfolgt.