

OCTRA BACKEND – EINE SKALIERBARE INFRASTRUKTUR FÜR TRANSKRIPTIONSPROJEKTE

Christoph Draxler, Julian Pömp

Institut für Phonetik und Sprachverarbeitung, LMU München
draxler@phonetik.uni-muenchen.de, j.poemp@phonetik.uni-muenchen.de

Kurzfassung: Octra Backend ist eine portable web-basierte Infrastruktur für Transkriptionsprojekte, die lokal im Feld oder geschützten Bereichen, im begrenzten Intranet oder weltweit erreichbar im Internet eingesetzt werden kann. Entwicklungsziele waren die Gewährleistung möglichst hoher Sicherheitsanforderungen, eine gute Skalierbarkeit sowie eine einfache Installation auch ohne Administratorrechte. Octra Backend ist in Node.js implementiert und für MacOS, Windows und Linux verfügbar.

1 Einführung

Das Erstellen von Sprachdatenbanken sowohl für die Technologieentwicklung als auch die Forschung ist auch in Zeiten großer Fortschritte in der KI-basierten Sprachverarbeitung eine zeitaufwendige und damit teure Aufgabe. Bereits 1997 wurden erste web-basierte Lösungen für die verteilte Transkription von Sprachaufnahmen präsentiert (z. B. WWWTranscribe von [1]) und in großen Sprachdatensammlungen ([2], [3]) eingesetzt, um den Ablauf der Transkription zu unterstützen und ein kollaboratives bzw. verteiltes Arbeiten zu ermöglichen. In der Regel erfordern diese Lösungen sowohl bei der Installation als auch im Betrieb Administratorrechte auf dem Server sowie Datenbank- und Administratorkenntnisse, so dass ihr Einsatz mit hohem Aufwand verbunden ist, der in Forschungsgruppen, Kliniken, Arztpraxen oder Firmen kaum zu leisten ist. Zudem verbieten der Schutz der Privatsphäre und Geheimhaltungsvorschriften, besonders in medizinischen oder Firmenumgebungen, häufig die Nutzung web-basierter Dienste durch externe Anbieter.

Octra Backend ist unseres Wissens nach die erste web-basierte Infrastruktur für Transkriptionsprojekte, die auf dem lokalen Rechner oder in geschlossenen lokalen Netzwerken, im Intranet sowie im Internet eingesetzt werden kann, die hohe Sicherheitsanforderungen unterstützt und sowohl ein grafisches als auch ein Programmierinterface anbietet.

Kapitel 2 gibt eine Übersicht über Octra Backend, Kapitel 3 beschreibt die Implementation, Kapitel 4 zeigt den Einsatz anhand eines Beispielprojekts, und Kapitel 5 fasst erste Erfahrungen zusammen und gibt einen Ausblick.

2 Octra Backend Übersicht

Octra Backend verwaltet die Erstellung von Sprachdatenbanken über Transkriptionsprojekte. Ein Transkriptionsprojekt besteht aus Audio- und zugeordneten Transkripten, Aufgaben, Transkriptionswerkzeugen sowie Accounts und Rollen mit jeweils unterschiedlichen Zuständigkeiten und Rechten.

2.1 Skalierbarkeit

Für Octra Backend sind drei Einsatzszenarien mit je typischen Netzkonfigurationen vorgesehen (siehe Abb. 1):

- a) Standalone-Installation für den Einsatz im Feld, in Praxen oder besonders geschützten Bereichen mit einem isolierten lokalen Netzwerk ohne Internetverbindung

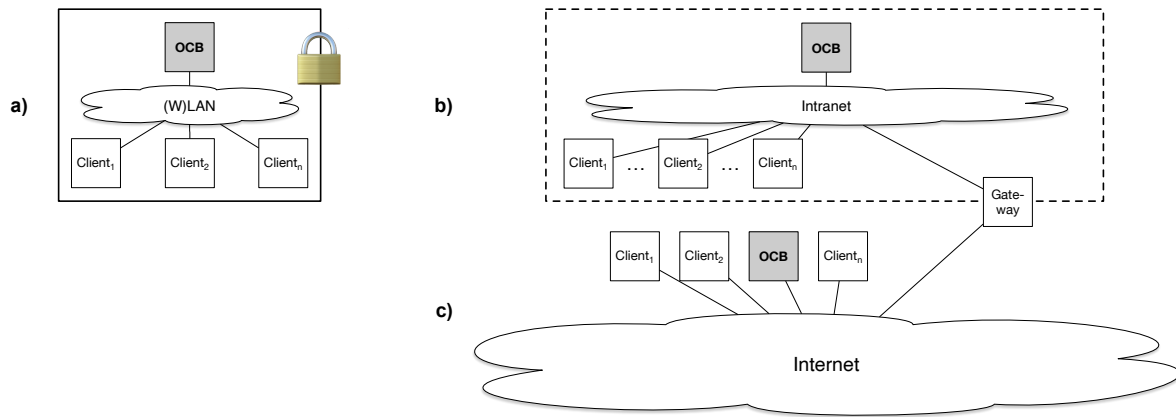


Abb. 1: Einsatzszenarien für Oetra Backend

- b) In Forschungs- und Arbeitsgruppen, Firmenabteilungen o.ä. mit einem Intranet mit einem Gateway ins Internet
- c) Als global erreichbarer Dienst im Internet

2.2 Sicherheit

Viele Sprachdatensammlungen stellen hohe Anforderungen an die Wahrung der Privatsphäre und der technischen Integrität – so ist es häufig nicht erlaubt, dass Daten auf Servern Dritter gespeichert und verarbeitet werden oder ‚das Haus verlassen‘ dürfen. Dies ist aus Sicht der Verarbeitung eine sehr große Einschränkung, denn damit werden bestimmte Verarbeitungsschritte ausgeschlossen, wie z.B. die Nutzung sprachverarbeitender Webdienste, oder sehr aufwendig, weil entweder der Dienst lokal installiert werden muss oder die benötigten Expertinnen und Experten nur vor Ort arbeiten dürfen. Diese Einschränkungen sind in der Regel darauf zurückzuführen, dass Daten an externe Dienstleister übertragen werden müssen und diese den Verarbeitungsweg nicht vollständig offenlegen oder sie die benötigten Sicherheitsanforderungen nicht gewährleisten können. Eine lokale Installation ist oft nicht möglich.

Diese grundsätzlichen Einschränkungen gelten auch für Oetra Backend. Allerdings bietet es durch seine Skalierbarkeit die Anpassung an die Erfordernisse verschiedener Einsatzszenarien, und innerhalb dieser Szenarien eine möglichst hohe Sicherheit bei der Kommunikation, der Datenverwaltung, der Abläufe und Werkzeuge sowie der Rechteverwaltung und Zugriffskontrolle:

2.3 Kommunikation

Im Einsatzszenario des geschlossenen Netzwerks läuft Oetra Backend auf einem dedizierten Server, z.B. einem Raspberry Pi, der ein eigenes lokales (W)LAN aufbaut. Physischen Zugang haben nur Endgeräte, die im lokalen Netzwerk bekannt sind. Eine verschlüsselte Kommunikation ist im lokalen Netzwerk nicht einfach möglich, weil dafür Zertifikate externer Anbieter notwendig sind¹. In den Szenarien Intra- und Internet ist die Kommunikation mit dem Backend verschlüsselt.

2.4 Datei- und Datenverwaltung

Oetra Backend legt die Audiodateien in einem eigenen Verzeichnis im lokalen Dateisystem ab, und jedes Projekt hat sein eigenes Projektverzeichnis. Dieses Verzeichnis ist geschützt und nur der Systemadministration zugänglich.

¹ Mit selbstgenerierten Zertifikaten ist die verschlüsselte Kommunikation auch in geschlossenen Netzwerken möglich. Eine Anleitung dafür findet sich in der Oetra Backend Hilfeseite.

Konfigurationen, Verwaltungsdaten, Dokumente und Transkripte sind Textdaten und werden in der Datenbank gespeichert.

Beim Import von Audio- und Transkriptdateien werden diese je nach Aufgabe miteinander verknüpft – z. B. Audiodateien mit von der Spracherkennung erstellten Rohtranskripten für die Korrektur; diese Verknüpfung kann automatisch erfolgen, etwa durch deckungsgleiche Dateinamen, oder manuell. Beim Import werden alle Dateien intern umbenannt, so dass sie auch von externen Webdiensten wie WebMAUS [4] o.ä. verarbeitet werden können, und es wird ein Hashwert für die Dateien berechnet, um zu verhindern, dass Dateien mehrfach geladen werden.

Interne Dateinamen und Pfade sind nach außen nicht sichtbar – die Werkzeuge arbeiten mit generierten Verweisen.

2.5 Accounts und Projektrollen

Accounts stehen für die Nutzer und Nutzerinnen des Backends. Sie bestehen aus einem frei wählbaren eindeutigen Namen sowie mindestens einer Emailadresse. Aus Gründen der Langfrist-Dokumentation werden Accounts typischerweise nicht gelöscht, sondern deaktiviert – auf diese Weise ist auch nach Projektende nachvollziehbar, wer wann welche Daten unter welchen rechtlichen Vereinbarungen bearbeitet hat.

Accounts können in einem Projekt verschiedene Rollen einnehmen, und diese Rollen können befristet werden. Auf Ebene des Systems sind die Rollen Systemadministration, Nutzer und Anwendung definiert, auf der Ebene der Projekte Projektadministration, Datenlieferant sowie Bearbeiter. Projektspezifisch können weitere Rollen definiert werden.

2.6 Aufgaben

Zentraler Aspekt der Transkriptionsverwaltung sind die Abläufe und Aufgaben. Eine Aufgabe umfasst den Aufgabentyp, z.B. Transkription, und die zu bearbeitenden Dateien, in der Regel eine Audiodatei und ein Rohtranskript. Die Zuordnung von Dateien zu Aufgaben erfolgt entweder beim Import der Dateien, oder zu einem späteren Zeitpunkt. Eine Aufgabe hat einen Status, z.B. frei, in Bearbeitung, zurückgestellt, abgeschlossen.

2.7 Transkriptionswerkzeuge

Aus Gründen der Sicherheit erlaubt Octra Backend nur den Einsatz bekannter Werkzeuge. Diese verwenden App-Tokens um sicherzustellen, dass eine Eingabe von einem autorisierten Werkzeug stammt. Aktuell unterstützt Octra Backend den Transkriptionseditor Octra [5].

Eine formale Überprüfung von Transkripten, z. B. durch reguläre Ausdrücke, wird unterstützt.

3 Implementation

Octra Backend ist ein Client-Server-System auf der Basis relationaler Datenbanken, das sowohl eine REST-Schnittstelle für den Programmzugriff als auch ein grafisches User Interface bietet. Für die Implementation wurde das Node.js Framework Nest.js verwendet, für das GUI das Angular Framework.

Als Datenbank kann prinzipiell jedes Datenbanksystem verwendet werden. Standardmäßig werden (www.sqlite.org/) sowie PostgreSQL (www.postgresql.org/) unterstützt. Abb. 2 zeigt eine schematische Darstellung der Datenbank.

Zur Installation sind drei Dateien notwendig: der Kommandozeileninterpreter, das Octra API sowie eine Konfigurationsdatei. Letztere enthält die Konfiguration in JSON-Format. Je

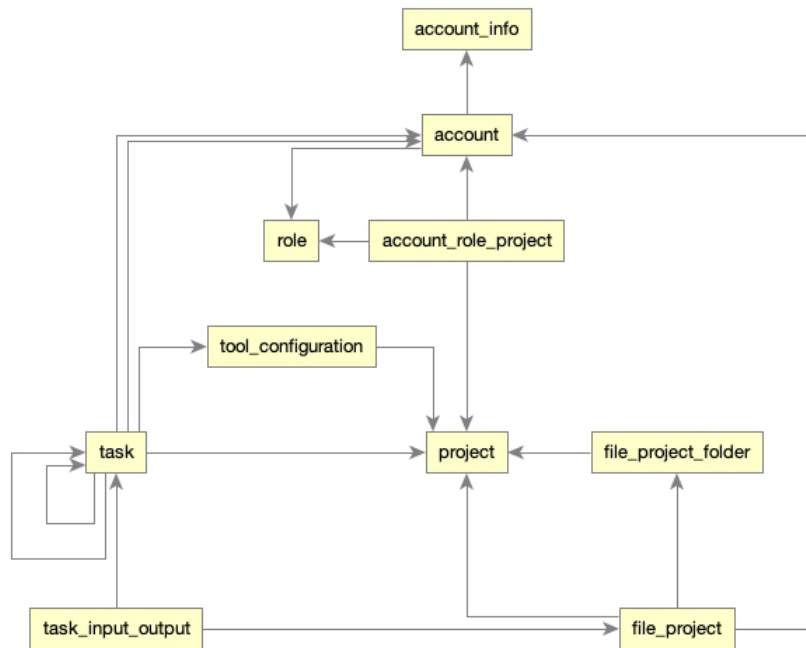


Abb. 2: schematisches Datenmodell

nach Einsatzszenario unterscheiden sich die zur Installation erforderlichen Rechte: für geschlossene Netze bzw. für ad hoc Arbeitsgruppen im Intranet ist eine Installation als einfacher Nutzer möglich, für das Internet sind in der Regel Administratorrechte notwendig.

Für die Installation sind für Windows, Linux (inkl. ARM 64 Bit) und MacOS eigene Installationskripte verfügbar, die eigentliche Software ist für alle Plattformen identisch.

Aktuell ist Octra Backend als privates Github-Repository verfügbar (github.com/IPS-LMU/octra-backend). Eine Veröffentlichung als Open-Source-Software ist geplant.

4 Beispielprojekt

In einem studentischen Projekt im Masterseminar Sprachdatenbanken am Institut für Phonetik der LMU im Wintersemester 2023/24 wurde eine Studie zur Veränderung der Sprechweise bei der Interaktion mit Haustieren [6] repliziert – Ablauf, Bildmaterial und Spielmaterial waren soweit möglich wie im Originalprojekt, nur waren die Haustiere hier Katzen (statt Hunde).

In diesem Projekt sind die Transkriptionsdaten im Octra Backend auf dem Server des Instituts installiert, so dass die Transkription auch von zuhause aus erfolgen kann.

Nachdem der Systemadministrator ein neues Projekt angelegt und einen Account zur Projektadministration angelegt hat, loggt sich diese ein und beginnt mit dem Hochladen der Daten. Dies erfolgt über eine Tabellenansicht, in die per Drag&Drop einzelne Dateien sowie Dateipaare, bestehend aus einem von einer Spracherkennung generierten Rohtranskript im .srt-Format sowie Audiodatei im .wav-Format eingefügt werden.

Nach dem Import stehen die Dateien für die Zuteilung zu Transkribiererinnen und Transkribierern zur Verfügung – diese Zuteilung ist in der Liste sichtbar. Je nach Dateityp sind Viewer zur Ansicht des Dateiinhalts verfügbar (Abb. 3)

Im nächsten Schritt legen die Transkribierer/innen ihre Accounts an; dazu werden sie aufgefordert, neben dem Nutzernamen und Email-Adresse auch Angaben zur weiteren Kenntnissen zu machen – dies sind z.B. Sprach- oder spezifische Transkriptionskenntnisse.

35. Konferenz Elektronische Sprachsignalverarbeitung

Anschließend weist die Projektadministration die Aufgaben zu: entweder werden die Aufgaben individuell einzelnen Personen zugewiesen, oder die Zuweisung bleibt offen, so dass ein/e Transkribierer/in stets die nächste freie Aufgabe bearbeitet.

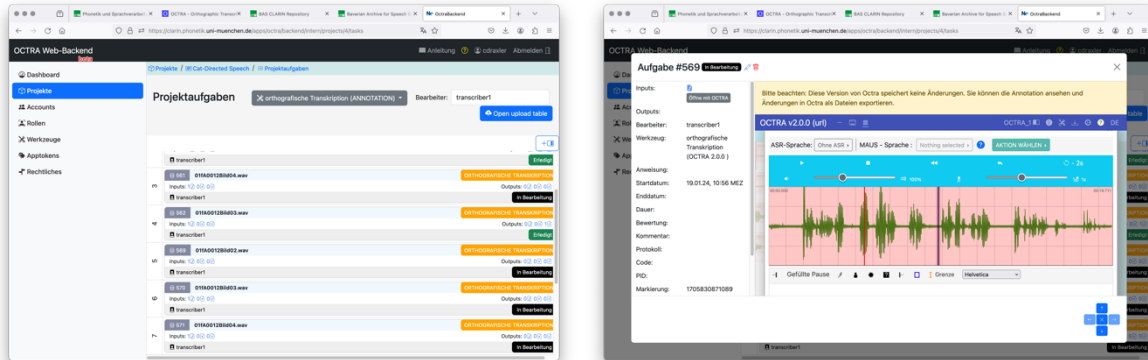


Abb. 3: Detailansichten von Aufgaben in Oetra Backend

5 Zusammenfassung und Ausblick

Oetra Backend ist noch in Entwicklung, kann aber bereits produktiv eingesetzt werden. Am IPS ist es in zwei komplementären Szenarien im Einsatz: 1) als allgemeiner Transkriptionsserver für verteilte Transkriptionsprojekte, auf der Basis der institutseigenen Server und PostgreSQL Datenbankinstallation sowie der Webdienste des BAS wie z.B. WebMAUS [4] und Spracherkennung, und 2) als autarke Installation auf einem Raspberry Pi; auf diesem Rechner sind auch der Editor Oetra, die Emu WebApp [7] zur Visualisierung installiert (Abb. 4). Geplant ist die Installation der Spracherkennung whisper von openAI (openai.com/research/whisper).

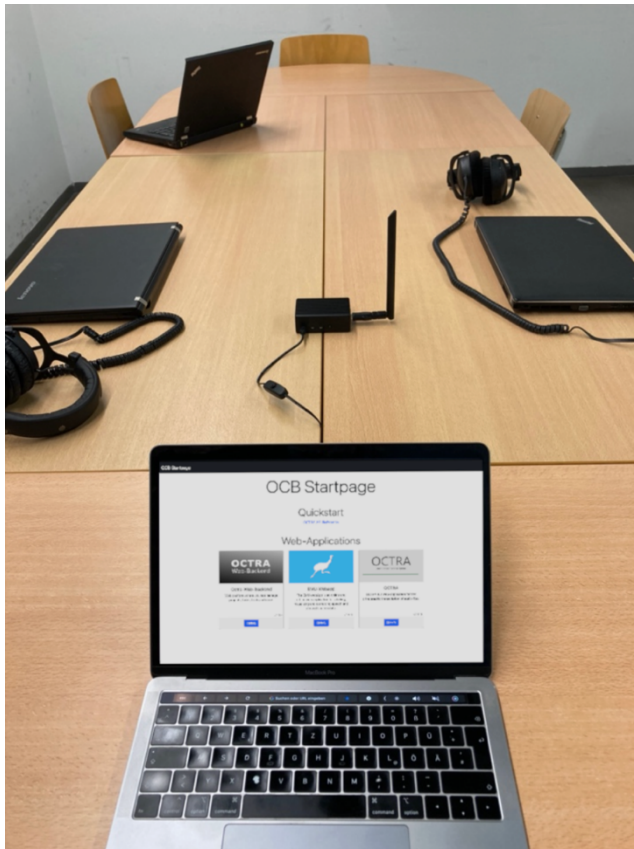


Abb. 4: Einsatz im Feld: Der Laptop vorne zeigt die Startseite der Software, auf dem Raspberry Pi (mitte) läuft Oetra Backend mit den Audio- und Transkriptionsdaten. Die Rechner sind im lokalen (W)LAN ohne Verbindung nach außen

Danksagung

Ein herzliches Danke geht an die Studierenden Camilla Anton, Henry Derwanz, Jana Freudenberger, Jonas Held, Magdalena Saumweber und Johanna Spateneder für die Idee und Durchführung des Projekts sowie an die Katzenhalterinnen.

Literatur

- [1] Draxler, Christoph (1997) WWWTranscribe – A Modular transcription system based on the World Wide Web, Proc. Eurospeech,, S. 1691–94.
- [2] Höge, Harald, Christoph Draxler, Henk van den Heuvel, Finn Johansen, Erik Sanders & Herbert Tropf (1999) SpeechDat Multilingual Speech Databases for Teleservices: Across the Finish Line, Proc. Eurospeech, S. 2699–702.
- [3] Draxler, Christoph & Klaus Jänsch (2006) Speech Recordings in Public Schools in Germany – the Perfect Show Case for Web-based Recordings and Annotation, in Proc. LREC, S. 2112–2115.
- [4] Kisler, Thomas, Florian Schiel & Han Sloetjes (2012) *Signal Processing Via Web Services: The Use Case WebMAUS*. in Proceedings Digital Humanities 30–34.
- [5] Draxler, Christoph & Julian Pömp (2022) *OCTRA – An Innovative Approach to Orthographic Transcription*. Proc. Interspeech, S. 5217-5218
- [6] Nuscheler, J.D, Elsässer, N., Braun, A. & Zahner-Ritter, K. (2023) *Acoustic characteristics of dog-directed speech: The role of individual empathy level and personality traits*. Proc. ICPHS
- [7] Winkelmann, Raphael & Georg Raess. (2014) Introducing a Web Application for Labeling, Visualizing Speech and Correcting Derived Speech Signals, Proc. LREC