

KI-basierte Content-Übersetzung in Leichte Sprache anhand von Webseiten öffentlicher Einrichtungen mit NLP-Modell von SUMM AI

Konstantin Kuklin

Bachelorarbeit • Studiengang Medieninformatik • Fachbereich Informatik und Medien • 23.08.2024

Aufgabenstellung

Ziel dieser Arbeit ist es, das deutsche NLP-Modell „SUMM AI“ zur automatisierten Übersetzung in Leichte Sprache in redaktionelle Prozesse öffentlicher Einrichtungen zu integrieren und dessen Effizienz zu evaluieren.

Konzept

Das ITZBund bietet Verwaltungen das CMS Government Site Builder an, das ab Juni 2024 auf TYPO3 basiert. Da die Version 11 zur Untersuchung noch nicht verfügbar war, wurde eine ältere TYPO3-Version genutzt (ITZBund Dienstsitz Bonn, 2024). Wegen der Auslagerung redaktioneller Prozesse an externe Dienstleister wurde ein Interview mit einer Agentur eines Bundesministeriums geführt, um den redaktionellen Ablauf zu verstehen. Für die Untersuchung wurden 12 Webtexte von drei Bundesministerien per Webcrawling erfasst, nach Inhaltstyp und Textlänge kategorisiert, und in einer nach dem Agentur-Ablauf eingerichteten TYPO3-Installation als Webseiten erstellt. Eine eigens entwickelte TYPO3-Erweiterung, basierend auf der TYPO3-Core API, übersetzte die Texte automatisch in Leichte Sprache mittels der SUMM AI API. Die Erweiterung fügte Felder für API-Token und Sprach-IDs hinzu und sendet bei CRUD-Operationen „new“ und „update“ automatisch POST-Requests, wenn Inhalte auf Deutsch vorliegen und noch keine Übersetzung in Leichte Sprache existiert.

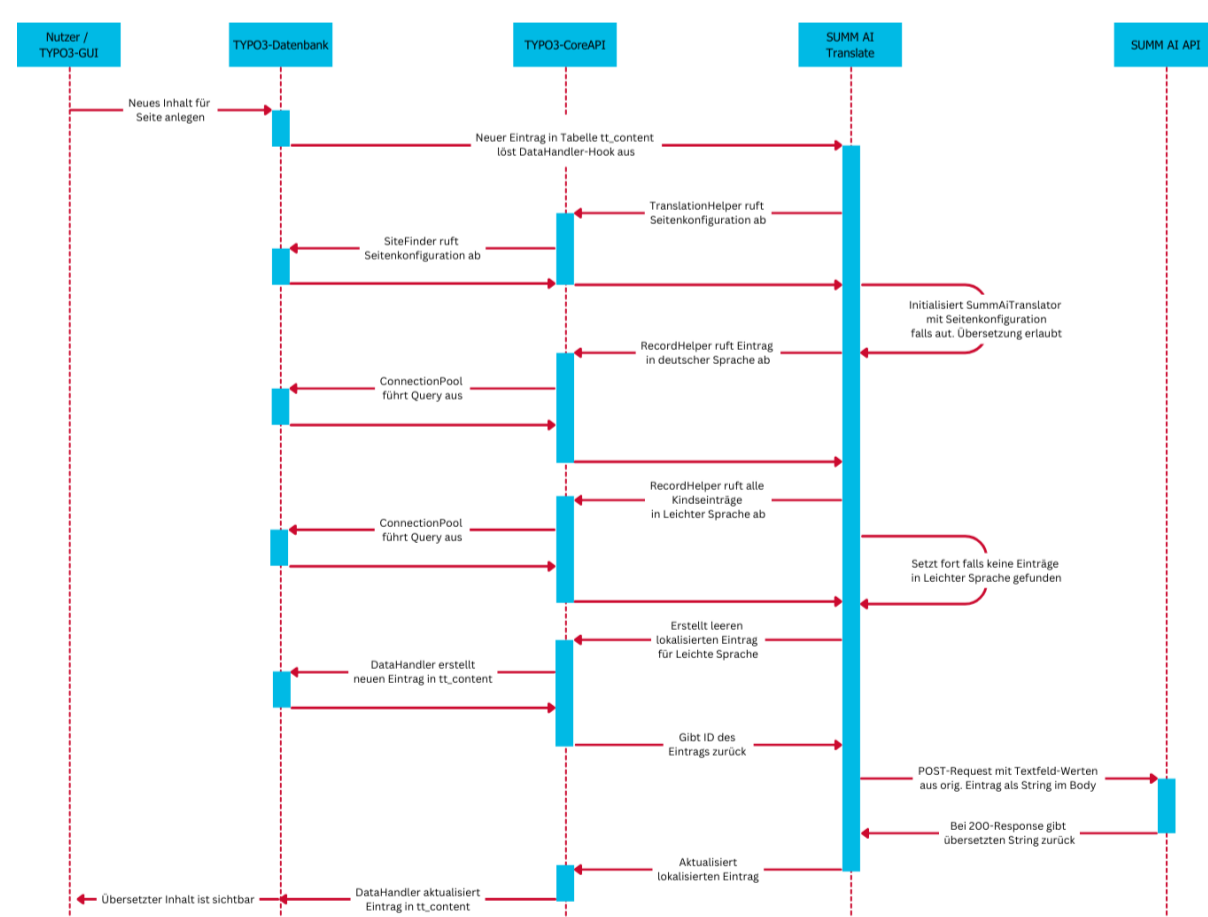


Abb. 1: Sequenzdiagramm für die TYPO3-Erweiterung

Leichte Sprache

BITV 2.0 verpflichtet öffentliche Verwaltungen, Informationen in Leichter Sprache bereitzustellen. Diese auf GER-Niveau A1-A2 verfassten Texte werden nach 13 festgelegten Regeln vereinfacht und zeichnen sich durch einfache Wörter, kurze Sätze, klare Struktur und den Verzicht auf Fachbegriffe aus, um sie für verschiedene Zielgruppen, darunter Sprachlernende und Menschen mit kognitiven Beeinträchtigungen, verständlich zu machen. Die übersetzten Textinhalte sind um 30% erhöhte Textlänge gekennzeichnet (Gross, 2023).

Testsystem

Um die Effizienz des NLP-Modells zu bewerten, werden die im Experiment erhobenen Daten analysiert, um drei Hypothesen zu prüfen. Die ersten beiden Hypothesen untersuchen die Abhängigkeit der Qualität von Textsorte und Textlänge, während die letzte Hypothese potenzielle Zeitersparnis misst. Als Qualitätskriterien dienen die formale Verständlichkeit, gemessen mit dem Hohenheimer Verständlichkeitsindex (zum Beispiel Satz- oder Wortlänge), die von Zielgruppenvertreter:in bewertete inhaltliche Korrektheit sowie die Einhaltung der BITV 2.0-Regeln.

Ergebnisse

Die Datenanalyse widerlegte zwei von drei Hypothesen. Während SUMM AI die Übersetzungszeit verkürzt, variiert die Qualität stark je nach Textsorte und -länge. Zwar verbessert das Modell die Lesbarkeit im Vergleich zu Standarddeutsch, jedoch leidet die Verständlichkeit aufgrund mangelnder Korrektheit und Regelkonformität, was die Texte für die Zielgruppe oft ungeeignet macht. Besonders „Informationsseiten“ und Texte über 3.600 Zeichen wiesen die geringste Qualität auf. SUMM AI steigert die Effizienz nur bei kurzen Texten unter 1.800 Zeichen; längere Texte erfordern signifikante Nachbearbeitung. Das Modell vereinfacht die Texte formal, etwa durch konsequente Zeilenumbrüche, aber eine wirkliche inhaltliche Vereinfachung wurde weder beim Abgleich des wiedergegebenen Inhalts mit den Kernaussagen noch bei der Überprüfung der Regeleinholung festgestellt.

Fazit

In dieser Arbeit wurde am Beispiel des CMS TYPO3 gezeigt, dass die Integration des NLP-Modells „SUMM AI“ für die Übersetzung in Leichte Sprache zwar möglich ist, die Qualität der Ergebnisse jedoch weiterhin menschliche Nachbearbeitung erfordert. Ergebnisse dieses NLP-Modells stimmen somit mit den in der Literaturrecherche betrachteten wissenschaftlichen Untersuchungen zu anderen Modellen, wie „ChatGPT“, überein.

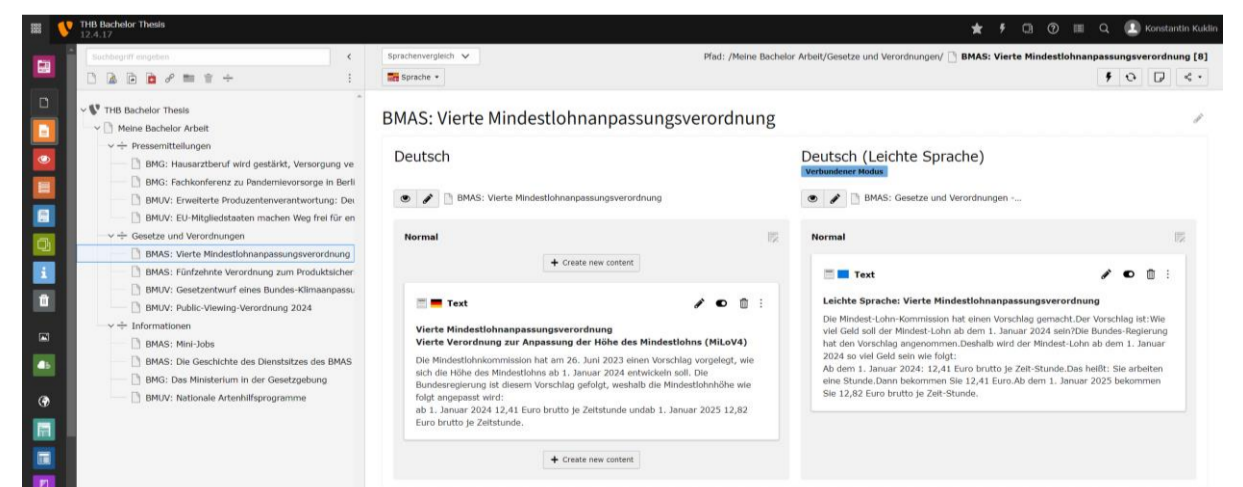


Abb. 2: Automatisch erzeugte Übersetzung durch SUMM AI

Quellen

- ITZBund Dienstsitz Bonn. (2024). *Government Site Builder 11*. Abgerufen am 2. August 2024 von Informationstechnikzentrum Bund: <https://produkt.gsb.bund.de/gsb11>
- Gross, S. (2023). *Leichte Sprache - Grundlagen und Anleitung für eine barrierefreie Kommunikation* (1. Ausg.). Bern: Hogrefe.