

Geoparsing vs. ChatGPT: Prototypische Ansätze zur Erkennung von Ortsbezügen aus Biographiegesprächen

Geoparsing vs. ChatGPT: Prototypical Approaches for Detecting Location References from Biographical Conversations

Alexander Rolwes, Patrick Stellbauer, Almut Plath, Klaus Böhm, Markus Schaffert, Sabine Corsten, Pascal Neis

Biographiearbeit zeigt das Potenzial, die Lebensqualität und das psychische Wohlbefinden bei Pflegebedürftigen signifikant zu verbessern. Dabei können besonders Orte Erinnerungen hervorrufen und positive Emotionen auslösen. Ziel dieser Forschungsarbeit ist es, vorbereitende technische Ansätze zu entwickeln, um Biographiegespräche hinsichtlich darin enthaltener Ortsbezüge automatisiert zu analysieren. Dazu werden ein Workflow implementiert, unterschiedliche lokale sowie online-basierte Erkennungsansätze analytisch gegenübergestellt und Erkenntnisse für den praktischen Einsatz abgeleitet. Es zeigt sich, dass der online-basierte Ansatz via ChatGPT zwar leichte Genauigkeitsvorsprünge gegenüber etablierten Geoparsing-Modellen aufweist, jedoch Schwächen in puncto Datenschutz und freier Erfindung von Orten zeigt.

Schlüsselwörter: Biographiearbeit, ChatGPT, Geoparsing, Ortsbezug

Biography work has the potential to significantly improve the quality of life and mental well-being of people in need of care. Places can evoke memories and trigger positive emotions. The aim of this research project is to develop preparatory technical approaches to automatically detect and identify place references in biography conversations. For this goal, we implemented a workflow, compare various local and online-based approaches analytically, and derive findings for practical use. The results show that the online-based approach via ChatGPT has a slight advantage in accuracy over established geoparsing-models, however, it shows weaknesses in terms of data protection and free invention of locations.

Keywords: *Biography work, ChatGPT, geoparsing, location reference*

1 EINLEITUNG UND MOTIVATION

Laut aktuellen Erhebungen waren im Jahr 2021 fünf Mio. Menschen in Deutschland im Sinne des Pflegeversicherungsgesetzes (SGB XI) pflegebedürftig, wovon insgesamt ca. 800 000 aller Pflegebedürftigen in Pflegeheimen vollstationär betreut wurden /Statista 2023/. Prognosen zufolge wird die Zahl der Pflegebedürftigen bis ins Jahr 2055 auf 6,8 Mio. Menschen ansteigen. Menschen im Alter sind oft durch beispielsweise Veränderungen des gewohnten Lebensumfelds

mit emotionaler und sozialer Einsamkeit, Anzeichen von Depressionen und einer reduzierten Lebensqualität konfrontiert, was sich negativ auf Gesundheit und psychisches Wohlbefinden auswirken kann /Drageset et al. 2011/, /Fraher & Coffey 2011/, /Luanaigh & Lawlor 2008/. Besonders im Alter führen Depressionen häufig zu begleitenden körperlichen Erkrankungen und nehmen einen chronischen Verlauf /Böhm et al. 2009/, was mit einer erhöhten

Sterblichkeitsrate einhergeht /Freudenstein et al. 2001/. Biographiearbeit ist ein etablierter Ansatz in der Altenpflege /Specht-Tomann 2018/, bei dem die individuelle Lebensgeschichte als Ausgangspunkt für die Verarbeitung der veränderten Lebenssituation und die Entwicklung eines positiven Identitätserlebens genutzt wird. Biographiearbeit hat das Potenzial, Anpassungsprozesse im Alter zu fördern und damit soziale Teilhabe und Lebensqualität zu unterstützen /Bauer et al. 2008/. Studien zeigen, dass Biographiearbeit signifikante Verbesserungen von Lebensqualität und psychischem Wohlbefinden bei älteren institutionalisierten Menschen mit geringen kognitiven Einschränkungen bewirken kann /Menn et al. 2020/. /Gennip et al. 2015/ identifizieren, dass Erinnerungen vor allem durch physische Objekte hervorgerufen werden (52 %), gefolgt von Orten (14 %), Aktivitäten (9 %), Menschen (9 %) und digitalen Items, wie Fotos oder soziale Medien (10 %). Orte können somit als Anker für Erinnerungen dienen und positive Emotionen auslösen /Gennip et al. 2015/. Allerdings können Orte nicht immer physisch aufgesucht werden. Eine Studie, in der durch virtuelle Reisen an Orten mit indirektem Raumbezug (z. B. Garten oder Kino) lebensgeschichtliche Erzählungen hervorgerufen werden, zeigt eine signifikante Verbesserung der Lebensqualität von Bewohnerinnen und Bewohnern in Senioreneinrichtungen /Corsten & Lauer 2020/, /Giordano et al. 2023/.

Dieser Ansatz wird in der vorliegenden Arbeit durch Fokus auf die Erkennung von direkten Ortsbezügen (Orte mit Koordinaten in einem Koordinatenreferenzsystem /ISO 19111 : 2019/) aus Biographiegesprächen weiter vertieft. Durch die gezielte Integration technikgestützter Biographiearbeit mit persönlich relevanten Orten soll die Lebensqualität von Bewohnern in Senioreneinrichtungen langfristig verbessert werden. Aktuelle Forschung berücksichtigt die Integration von Ortsbezügen mit direktem Raumbezug in Biographiegesprächen jedoch nicht. Daher stellt sich die Forschungsfrage, wie und in welcher Qualität Ortsbezüge mit direktem Raumbezug aus Biographiegesprächen automatisiert erkannt werden können. Zur Beantwortung dieser Frage wird eine Untersuchung durchgeführt, die lokale Geoparsing-Ansätze mit einem online-basierten ChatGPT-Ansatz (Generative Pre-trained Transformer (GPT)) zur Erkennung von Ortsbezügen analytisch evaluiert und Erkenntnisse für den praktischen Einsatz ableitet. Angesichts der Aktualität und frühen Entwicklungsstufe von ChatGPT besteht ein Forschungsbedarf bezüglich der Unterschiede aufstrebender Chatbots zu etablierten Geoparsing-Modellen. Die Arbeit implementiert einen funktionsfähigen Workflow, der audiovisuell aufgezeichnete Biographiegespräche in verschiedenen Prozessschritten hinsichtlich enthaltener Ortsbezüge auswerten kann.

2 STAND DER TECHNIK

Seit den Anfängen in den 1990er-Jahren sind Modelle zur Umwandlung von Sprache zu Text (Speech-to-Text) Teil der Forschung /Kurzweil 1999/. Damals basierten Spracherkennungsalgorithmen primär auf Hidden-Markov-Modellen und lieferten ausreichende, wenn auch ungenaue Ergebnisse in extrahierten Transkripten /Rabiner 1989/. Im Laufe der letzten Jahrzehnte hat die Forschung sowie der Fortschritt in der künstlichen Intelligenz (KI) dazu beigetra-

gen, Spracherkennungsalgorithmen in alltägliche Anwendungsfelder und auf unterschiedlichen Endgeräten einzubinden. Es existiert mittlerweile eine Vielzahl an kostenpflichtiger und kostenfreier Spracherkennungssoftware, welche unterschiedliche Schwerpunkte und Charakteristiken aufweisen /Potamianos et al. 2004/, /Benzeghiba et al. 2007/, /Arora & Singh 2012/, /Yu & Deng 2015/, /Alharbi et al. 2021/. Als Evaluationsmetrik zur praktischen Leistungsbewertung der Spracherkennung finden die Word Error Rate (WER), die Word Recognition Rate (WRR) und die Sentence Error Rate (SER) Anwendung /Hunt 1990/, /Klawow & Peters 2002/. Da sich diese Metriken ausschließlich auf Wörter und deren Erkennungsqualität fokussieren, wird in diesem Beitrag zusätzlich eine neue, ortsbezogene Bewertung (nachfolgend bezeichnet als Word Location Error Rate (WLER)) geschaffen und eingeführt. Diese prüft, wie viele der im Text befindlichen Orte tatsächlich fehlerfrei transkribiert werden, um Rückschlüsse auf mögliche Fehlkoordinaten zu schließen.

Zur Extraktion und Koordinierung räumlicher Informationen aus Texten existiert eine Vielzahl an sogenannten Geoparsing-Verfahren /Gregory et al. 2015/, /Hu et al. 2022/. Diese verwenden unterschiedliche Modelle des Natural Language Processing (NLP), um Ortsnamen in Texten in einem zweistufigen Prozess zu identifizieren (Geoannotation) und in Koordinaten umzuwandeln (Geocoding). Diese Modelle sind ein Teilbereich der KI, welche es Computern ermöglicht, Sprache auf ähnliche Weise wie Menschen zu verstehen, zu verarbeiten und zu generieren /Cowie & Lehnert 1996/. Im ersten Schritt der Geoannotation werden im Toponym-Auflösungsverfahren Freitextbeschreibungen von Orten (Entitäten) in Texten identifiziert. Dabei konzentriert sich die Wahl der Verfahren in der Regel auf Named-Entity-Recognition-(NER-)Modelle mit Bibliotheken wie Spacy, Flair oder Stanford /Gritta et al. 2020/, /Hu et al. 2022/. Die NER-Technologie wird gegenwärtig mit unterschiedlichen Auswertungsansätzen (regelbasiert, gazetterbasiert, statistische oder hybride Ansätze) in verschiedenen Anwendungsbereichen eingesetzt, darunter beispielsweise maschinelle Übersetzungen, Analyseunterstützung im Katastrophenmanagement oder Social-Media-Analysen /Nadeau & Sekine 2007/, /Ruthmann et al. 2020/. Bei der Auswahl und Anwendung von NER-Software müssen u. a. die Leistungsfähigkeit, Kosten, Dokumentation, Datenschutz oder Lizenzierung berücksichtigt werden. Dabei können NER-Modelle neben Entitäten zu Orten auch personen- (beispielsweise Personennamen) oder organisationsbezogene (beispielsweise Unternehmen oder Sportvereine) Entitäten identifizieren. /Purves et al. 2018/ präsentierten ein Schema zur Bewertung der Leistungsfähigkeit von Verfahren zur Geoannotation (vgl. Tab. 1). Ein zentrales Element dabei ist ein manuell erstellter Referenzdatensatz, mit dessen Vergleich zu den vom Geoannotationsmodell markierten Toponymen sich alle korrekt identifizierten Toponyme (True Positive (TP)), alle fehlerhaften Markierungen (False Positive (FP)) und alle übersehenen Toponyme (False Negative (FN)) ableiten lassen. Diese Werte bilden die Grundlage zur Berechnung der Gütekriterien Precision, Recall und F1-Score, um einen Vergleich zwischen verschiedenen Auflösungsverfahren zu ermöglichen /Karimzadeh et al. 2019/, /Gritta et al. 2020/.

Zur Bewertung der Klassifizierungsmodelle gilt der F1-Score als anerkannte Metrik, da dieser ein Gleichgewicht zwischen Precision

	Anwendungsfallbezogene Definition	Mathematische Gleichung
True Positive (TP)	Positive Klasse richtig vorhergesagt. Fälle, in denen Location-Entitäten korrekt als Orte identifiziert wurden.	—
False Positive (FP)	Positive Klasse falsch vorhergesagt. Fälle, in denen Non-Location-Entitäten als Orte identifiziert wurden.	—
False Negative (FN)	Negative Klasse falsch vorhergesagt. Fälle, in denen „echte“ Orte nicht als Location-Entitäten identifiziert wurden.	—
Precision	Der Anteil der Entitäten, die korrekt als Orte identifiziert wurden (TP).	$\text{Precision} = \frac{TP}{TP + FP}$
Recall	Der Anteil aller identifizierten Orte (TP) zeigt, wie gut das Modell positive Ergebnisse identifiziert.	$\text{Recall} = \frac{TP}{TP + FN}$
F1-Score	Gleichgewicht zwischen Precision und Recall zur Leistungsmessung des Modells.	$\text{F1-Score} = W[0;1] = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$

Tab. 1 | Anwendungsfallbezogene Definition und mathematische Gleichungen zur Bewertung der Leistungsfähigkeit von Verfahren zur Geoannotation nach /Purves et al. 2018/, /Karimzadeh et al. 2019/, /Gritta et al. 2020/, /Sesselmann et al. 2023/

und Recall herstellt und eine alleinige Punktzahl zum Vergleich von Modellen bietet. Nach einer umfassenden Untersuchung existierender NER-Modelle ist Flair derzeit das Modell mit der höchsten Precision und einem hohen F1-Score /Hu et al. 2022/. Gleichzeitig kann die Trennung zwischen den Entitätsklassen Ort und Organisation unscharf sein, was wiederum teilweise in niedrigen F1-Scores anderer Modelle resultiert. Auch im hier vorliegenden Untersuchungsgegenstand wird die Evaluation über einen Typonymvergleich mit einem manuell erstellten Referenzdatensatz angestrebt.

Um anschließend im zweiten Schritt des Geoparsings die identifizierten Orte in eindeutige geographische Kennungen zu überführen, existieren unterschiedliche Geocoding-Ansätze. Dabei wird beispielsweise einer Adresse oder einer Sehenswürdigkeit die entsprechende geographische Koordinate zugewiesen, um detaillierte räumliche Analysen zu ermöglichen. Dies erlaubt z. B. die Identifikation von Mustern, Trends und Beziehungen zwischen verschiedenen Standorten. Darüber hinaus bietet das Geocoding die Möglichkeit, identifizierte Entitäten nutzerorientiert auf interaktiven Karten zu visualisieren. Als anerkannte Open-Source-Bibliothek wird u. a. Nominatim identifiziert /Nominatim developer community o. J./.

Vorteil dieser Bibliothek ist eine datenschutzkonforme, lokale Verwendungsmöglichkeit. Die Genauigkeitsbewertung der resultierenden Koordinaten geschieht über einen Koordinatenvergleich.

Neben dem zuvor erläuterten Prozess des Erkennens von Ortsbezügen existiert spätestens durch den im Jahr 2022/2023 entstandenen „ChatGPT-Boom“ eine moderne, vereinfachte Möglichkeit zur Verwendung von KI /Metoyer et al. 2018/, /Kano et al. 2019/. ChatGPT ist ein von OpenAI entwickelter Chatbot, welcher moderne maschinelle Lerntechnologien verwendet, um menschenähnliche Antworten zu generieren /OpenAI 2023/. Die Wissensbasis von ChatGPT erstreckt sich über einen umfassenden Textkorporus aus Büchern, Wikipedia-Einträgen, Webseiten oder auch literarischen Textsammlungen, ohne jedoch genaue Informationen zu den Trainingsdaten zu veröffentlichen. Dieser Informationsschatz kann unserer Meinung nach auch zum Geoparsing von Ortsbezügen in Texten genutzt werden. Erste Forschungsergebnisse zur Identifikation von Standortdaten aus Social-Media-Beiträgen mit GPT-

Modellen /Serere et al. 2023/, /Hu et al. 2023/ unterstützen diese Meinung.

3 METHODIK

Die Untersuchung im Rahmen dieses Beitrags sieht ein dreistufiges Vorgehen mit jeweils unterschiedlichen Testfällen vor. Ziel ist es dabei einerseits, unterschiedliche Erfahrungen im Umgang mit den verschiedenen Prozessschritten zu sammeln, um schlussendlich einen funktionsfähigen Workflow zur automatisierten Erkennung von Ortsbezügen aus Biographiegesprächen bereitzustellen. Die erste Untersuchung erfolgt dabei unter laborähnlichen Bedingungen. Hierzu wurde eine von KI erzeugte, textliche Lebensreise durch eine KI vorgelesen und audiovisuell aufgezeichnet, sodass keinerlei Hintergrundgeräusche, fehlerhafte Aussprache oder Dialekte diese Aufnahme verfälschen. Hieran wurden primär der Workflow entwickelt, verschiedene Parameter justiert sowie erste Erkenntnisse abgeleitet. Die zweite Untersuchung legt den Fokus auf eine erhöhte Anzahl an Ortsbezeichnungen in einem Video. Dazu wurde ein Youtube-Reisevideo über Wien ausgewertet und hinsichtlich der zuvor abgeleiteten Parameter verfeinert. Gleichzeitig existieren in diesem Reisevideo aufgrund der Aufnahme vor Ort einige Hintergrundgeräusche, wodurch Ergebnisse ggf. verfälscht werden können. Die abschließende dritte Untersuchung analysiert einen Testfall ähnlich zu einem echten Biographiegespräch. Da zum Zeitpunkt der Untersuchungen das Forschungsvorhaben einer Prüfung durch eine Ethikkommission zum Schutz der vulnerablen Zielgruppe unterliegt, kann gegenwärtig nicht mit echten Gesprächsdaten gearbeitet werden. Daher wurde ein fiktives, adaptiertes Biographiegespräch mit Fokus auf eine erhöhte Anzahl an Ortsbenennungen in einem projektinternen Dialog zwischen einer Sprachtherapeutin und einem Datenwissenschaftler durchgeführt, um eine maximale Ähnlichkeit zu einem realen Gespräch sicherzustellen. Dieses Vorgehen profitiert zudem vom mehrjährigen Erfahrungsschatz in der Biographiearbeit.

Zur Beantwortung der Forschungsfrage wird ein prototypischer Workflow entwickelt, welcher entlang verschiedener Prozessschritte

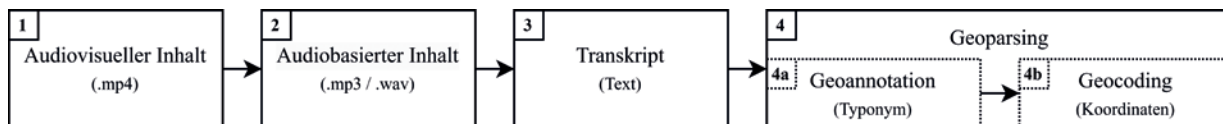


Abb. 1 | Technisches Konzept zur automatisierten Erkennung von Ortsbezügen in Biographiegesprächen

automatisiert Ortsbezüge aus Biographiegesprächen erkennt (Abb. 1). Dieser Workflow wird in sowohl lokal und online-basierten Verfahren für das Transkript (3) als auch Geoparsing (4) umgesetzt. Dabei besteht das Geoparsing aus den Teilen Geoannotation (4a) (Identifikation von Ortsentitäten) und Geocoding (4b) (Überführung der erkannten Orte in Koordinaten).

Als Eingangsmedium des Workflows dienen audiovisuelle Aufnahmen (1) der Gespräche. Diese audiovisuellen Aufnahmen werden zunächst in einen audiobasierten Inhalt (2) extrahiert und anschließend transkribiert (3), sodass zum Geoparsing (4) nur textbasierte Inhalte verwendet werden. Eine transkribierte Form der Gespräche erleichtert einerseits die nachfolgende Analyse, andererseits kann sie zu Dokumentationszwecken und zur therapeutischen Wiederaufnahme in der Biographiearbeit dienen. Zur Konvertierung der audiovisuellen in audiobasierte Inhalte (Export des akustischen Signals) existieren verschiedene kostenfreie sowie kostenpflichtige Bibliotheken. Im Zuge der Untersuchung hat sich die etablierte Open-Source-Python-Bibliothek *pydub* als geeignete Lösung herausgestellt /Porwal et al. 2022/. Vorteil dieser ist die lokale Datenverarbeitung, die maximale Zuverlässigkeit des Ergebnisexports in allen vorliegenden Anwendungsfällen sowie die Möglichkeit, Inhalte zu teilen und Störgeräusche zu reduzieren. Der Schwerpunkt der Untersuchungen im Rahmen dieses Beitrags liegt auf den nachfolgenden zwei Prozessschritten: Erzeugung eines Transkripts aus audiobasierten Inhalten (3) sowie die anschließende Identifikation und Koordinierung von Ortsbezügen mittels Geoparsing (4). Dazu wird eine Unterscheidung zwischen lokalen und online-basierten Analyseverfahren in beiden Schritten angestrebt, um die jeweiligen Stärken und Schwächen zu identifizieren sowie eine Vergleichbarkeit für die Untersuchung zu gewährleisten. Wie in Abschnitt 2 erläutert, existiert eine Vielzahl an verschiedener Spracherkennungssoftware zur Umsetzung von Prozessschritt 3. Während der laufenden Untersuchung fiel die Wahl auf die anerkannte Open-Source-Software *Whisper* von OpenAI /Radford et al. 2022/, welche in Python nutzerfreundlich zu integrieren ist und gleichzeitig eine lokale sowie online-basierte Umsetzung gewährleistet.

Wie eingangs beschrieben, dient das resultierende Transkript nachfolgend zum Geoparsing. Auch in diesem Prozessschritt wird zwischen einer lokalen Variante via Flair und Nominatim sowie einem online-basierten Vorgehen via ChatGPT-Programmierschnittstelle (engl. Application Programming Interface (API)) unterschieden. Die Besonderheit hierbei ist, dass die lokale Open-Source-Implementierung für Geoannotation (4a) sowie Geocoding (4b) unterschiedliche Softwarebibliotheken nutzt, wohingegen die online-basierte Implementierung beide Schritte vereint. Daraus resultieren zudem unterschiedliche Parameter, mit welcher die jeweiligen Ansätze feinjustiert werden können. Besonders hervorzuheben ist, dass

prompt content = „Extrahiere alle Orte aus dem Text und verknüpfe sie mit passenden Koordinaten. Die Koordinaten sollen im folgenden Muster ausgegeben werden: 52.5200,13.4050. Gib die Locations zusammen mit den Koordinaten als CSV aus. Verwende drei Spalten: 'Location, lat, lon'.“

temperature = 0.1 und 0.5
max_tokens = 2000 und 3000

Befehl 1 | Verwendeter Prompt und Modellparameter für die ChatGPT API

die durchgeführte lokale Geoannotation den transkribierten Text nur hinsichtlich ortbezogener Entitätsklassen (LOC, location name) untersucht. Organisationsbezogene Entitäten (ORG, organization name) werden wegen der erwähnten Unschärfe in der Trennung zwischen beiden Entitätsklassen nicht berücksichtigt. Dadurch wird die Anzahl der FP-klassifizierten Orte verringert. Die Evaluation der Geoannotation wird über einen Typonymvergleich mit einem manuell erstellten Referenzdatensatz durchgeführt.

Im Geocoding via Nominatim wurde eine eigene Instanz verwendet. Der alternative ChatGPT-basierte API-Ansatz verwendet das aktuell kostenfreie Modell „gpt-3.5-turbo“. Als Übergabenachricht wurde der API der transkribierte Text, der Prompt und die Modellparameter aus *Befehl 1* übermittelt. Bei den Parametern *temperature* und *max_tokens* wurde bewusst zwischen zwei Werten unterschieden. Der Parameter *temperature* beeinflusst die Zufälligkeit der generierten Antworten, sodass höhere Werte die Antwort vielfältiger, aber gleichzeitig auch freier gestalten, niedrigere Werte hingegen die Antwort explizierter und deterministischer ausführen /OpenAI Platform 2023/. Für die automatisierte Identifikation der Orte lässt sich daher steuern, wie fokussiert Orte erkannt werden sollen, ohne erfundene Orte oder Koordinaten mit einzubeziehen. Der Parameter *max_tokens* begrenzt die Länge der generierten Antworten /OpenAI Platform 2023/. Hier gilt es aufgrund des Token-basierten Kostenmodells einerseits einen plausiblen Kosten-Nutzen-Faktor zu identifizieren, gleichzeitig aber dem Chatbot auch den Raum zu gewähren, eine möglichst hohe Anzahl an Ortsbezügen mit zugehörigen Koordinaten zu erkennen.

	Transkript	Geoparsing	
		Geoannotation	Geocodierung
Trans _{Local} -Geo _{Online_lowParams}	Whisper _{Local}	ChatGPT _(max_tokens = 2000; temperature = 0.1)	
Trans _{Local} -Geo _{Online_highParams}	Whisper _{Local}	ChatGPT _(max_tokens = 3000; temperature = 0.5)	
Trans _{Online} -Geo _{Online_lowParams}	Whisper _{Online}	ChatGPT _(max_tokens = 2000; temperature = 0.1)	
Trans _{Online} -Geo _{Online_highParams}	Whisper _{Online}	ChatGPT _(max_tokens = 3000; temperature = 0.5)	
Trans _{Online} -Geo _{Local}	Whisper _{Online}	Flair	Nominatim
Trans _{Local} -Geo _{Local}	Whisper _{Local}	Flair	Nominatim

Tab. 2 | Kombinationsmöglichkeiten in der Untersuchung „Fiktives Biographiegespräch“

Aufgrund der genannten Unterscheidungen zwischen dem lokalen und dem online-basierten Ansatz sowie der verschiedenen ChatGPT-basierten Modellparameter ergeben sich in den finalen Untersuchungen zum Testfall des fiktiven Biographiegesprächs sechs Kombinationsmöglichkeiten, welche tiefergehend untersucht werden (Tab. 2).

4 UNTERSUCHUNG

Zur konkreten Untersuchung des lokalen und des online-basierten Ansatzes werden nachfolgend zwei Testfälle mit unterschiedlichen audiovisuellen Eingangsmedien zur Erkennung von Ortsbezügen analysiert. Dabei werden diese hinsichtlich ihrer Genauigkeit in den einzelnen Prozessschritten evaluiert und ein mögliches Optimierungspotenzial ermittelt.

4.1 Reisevideo

Der Testfall „Reisevideo“ führt eine vollständige Analyse sowie eine Genauigkeitsbetrachtung der automatisierten Detektion von Ortsbezügen aus audiovisuellen Inhalten durch. Zur Durchführung wurde ein Video recherchiert, welches einem realen Anwendungsfall ähnelt und eine große Anzahl an Ortsbenennungen beinhaltet. Kurze Reisevideos bieten hierzu eine optimale Grundlage, da sie viele unterschiedliche Ortsangaben enthalten und gleichzeitig ggf. durch Hintergrundgeräusche, undeutliche Aussprache oder die Redensweise von Dialekten Herausforderungen für den entwickelten Workflow darstellen. Gegenstand des Testfalls ist das Video „Wien in 5 Minuten | Reiseführer | Die besten Sehenswürdigkeiten“ des Kanals „theTravellers“ auf dem Videoportal YouTube /theTravellers 2016/. Das Video ist in deutscher Sprache aufgezeichnet und beinhaltet insgesamt 42 manuell erhobene Ortsangaben.

Tab. 3 zeigt die Evaluationsergebnisse des Prozessschritts *Transkript* für sowohl die lokale als auch die online-basierte Instanz. Die Ergebnisse zeigen, dass die online-basierte Instanz (Trans_{Online}) für alle Parameter eine höhere Genauigkeit gegenüber der lokalen Instanz (Trans_{Local}) aufweist. Insbesondere bei Betrachtung der

WER sowie der WRR zeigt sich die Überlegenheit von Whisper_{Online} im Vergleich zu Whisper_{Local}. Wie bereits in Abschnitt 2 erläutert, wird in diesem Beitrag der primäre Fokus auf die Erkennung von Orten gelegt, weshalb zusätzlich die WLER analysiert wird. Hier werden in der online-basierten Instanz 9,5 Prozentpunkte mehr Ortsangaben korrekt transkribiert, was gleichbedeutend mit der Anzahl von vier Orten ist. Fehler im Transkript können sich auf Fehlzuordnungen oder Fehlidentifikationen in den nachfolgenden Schritten auswirken. In diesem Testfall wird aufgrund dieser Ergebnisse nachfolgend nur die online-basierte Instanz (Whisper_{Online}) für den Prozessschritt *Transkript* weiterverwendet.

Der Prozessschritt *Geoannotation* bestimmt die ortbezogenen Entitäten aus dem zuvor online-transkribierten Text. In diesem Prozessschritt zielt die Untersuchung darauf ab, eine lokale Umsetzung mit Flair_(GeoLocal) gegenüber einer online-basierten Umsetzung mit ChatGPT_(GeoOnline_lowParams) zu vergleichen. Tab. 4 zeigt die Evaluationsergebnisse dieses Prozessschritts für beide Ansätze. Um die Genauigkeit des Geoannotation separat zu betrachten, bezieht sich die Berechnung der Gütekriterien auf die korrekt transkribierten Orte ($n = 38$) aus dem vorherigen Prozessschritt (vgl. Tab. 3). Es zeigt sich für den Prozessschritt *Geoannotation*, dass die lokale Umsetzung eine deutlich höhere Anzahl an identifizierten ortsbezogenen Entitäten gegenüber der online-basierten Umsetzung aufweist (28 TP vs. 20 TP). Gleiches bestätigt der um etwa 14 Prozentpunkte höhere F1-Score aufseiten der lokalen Umsetzung mit Flair. Als Grund für die hohe Anzahl an FN aufseiten der online-basierten Umsetzung lässt sich benennen, dass ChatGPT Schwächen in der Erkennung mehrfach genannter Orte trotz unterschiedlicher Prompts zeigt. Konkret bedeutet das, dass obwohl Orte im Reisevideo mehrfach genannt werden, diese nur einmal als Entität ausgegeben werden. Aus diesem Grund wird zusätzlich in Tab. 4 (siehe Zeile *) das mögliche Potenzial von ChatGPT aufgelistet, insofern diese Problematik gelöst wird. Es zeigt sich, dass die Anzahl der TP dann auf 39 ansteigen kann, was einen F1-Score von 0,951 ausmacht und somit deutlich über dem der lokalen Umsetzung mit Flair (F1-Score = 0,812) liegt.

Der Prozessschritt *Geocoding* überführt abschließend die zuvor identifizierten ortsbezogenen Entitäten in Koordinaten. Auch in diesem Prozessschritt wird eine Unterscheidung von lokaler zu online-

	WER (Word Error Rate)	WRR (Word Recognition Rate)	SRR (Sentence Error Rate)	WLER (Word Location Error Rate)
Trans _{Online}	3,8 % (18/471)	97,2 % (458/471)	100 % (1/1)	90,5 % (38/42)
Trans _{Local}	10,6 % (50/471)	90,9 % (428/471)	100 % (1/1)	81,0 % (34/42)

Tab. 3 | Evaluationsergebnisse des Prozessschritts *Transkript* in der Untersuchung „Reisevideo“

	TP (True Positive)	FP (False Positive)	FN (False Negative)	Precision	Recall	F1-Score
c) Trans _{Online} -Geo _{Online_lowParams}	20	1	18	0,952	0,526	0,678
*	39	1	3	0,975	0,929	0,951
e) Trans _{Online} -Geo _{Local}	28	4	9	0,875	0,757	0,812

* Mögliches Optimierungspotenzial, wenn ChatGPT mehrfach genannte Orte mehrfach identifizieren würde.

Tab. 4 | Evaluationsergebnisse des Prozessschritts *Geoannotation* in der Untersuchung „Reisevideo“

	Korrekt identifizierte und geocodierte Orte	
c) Trans _{Online} -Geo _{Online_lowParams}	67,5 %	(13,5/20)
*	80,8 %	(31,5/39)
e) Trans _{Online} -Geo _{Local}	53,6 %	(15/28)

* Mögliches Optimierungspotenzial, wenn ChatGPT mehrfach genannte Orte mehrfach identifizieren würde.

Tab. 5 | Evaluationsergebnisse des Prozessschritts *Geocoding* in der Untersuchung „Reisevideo“

	Korrekt identifizierte und geocodierte Orte	
c) Trans _{Online} -Geo _{Online_lowParams}	32,1 %	(13,5/42)
*	75,0 %	(31,5/42)
e) Trans _{Online} -Geo _{Local}	35,7 %	(15/42)

* Mögliches Optimierungspotenzial, wenn ChatGPT mehrfach genannte Orte mehrfach identifizieren würde.

Tab. 6 | Evaluationsergebnisse des gesamten Workflows in der Untersuchung „Reisevideo“

basierter Umsetzung dargestellt. *Tab. 5* zeigt die Evaluationsergebnisse dieses Prozessschritts für beide Ansätze. Um die Genauigkeit des Geocodings separat zu betrachten, bezieht sich die Berechnung der korrekt geocodierten Orte hier auf die identifizierten ortsbezogenen Entitäten ($n = 20$ bzw. $n = 39$ bzw. $n = 28$) aus dem vorherigen Prozessschritt (vgl. *Tab. 4*). Die Ergebnisse zeigen, dass via ChatGPT der größte prozentuale Anteil (67,5 %) an korrekt geocodierten Orten bezogen auf die zuvor ermittelten Entitäten erreicht wird, auch wenn die absolute Anzahl (13,5/20 korrekt geocodierte Orte) geringer ist als bei der lokalen Umsetzung (53,6 % $\hat{=}$ 15/28 korrekt geocodierte Orte) via Nominatim. Auch in diesem Prozessschritt wird das mögliche Potenzial von ChatGPT aufgezeigt, insofern eine Mehrfacherkennung von genannten Orten stattfindet. Es zeigt sich dann ein Optimierungspotenzial auf 80,8 % $\hat{=}$ 31,5/39 korrekt geocodierter Orte (siehe Zeile *).

Tab. 6 zeigt die Evaluationsergebnisse des gesamten Workflows für beide Ansätze. Der lokale Ansatz mit Flair und Nominatim konnte eine leicht höhere Anzahl (35,7 % $\hat{=}$ 15/42) an korrekt identifizierten und geokoordinierten Orten gegenüber dem online-basierten Ansatz mit ChatGPT (32,1 % $\hat{=}$ 13,5/42) aufweisen. Gleichzeitig zeigt sich auch hier ein deutliches Optimierungspotenzial (siehe Zeile *) mit ChatGPT (75,0 % $\hat{=}$ 31,5/42) bei einer Behebung der Schwäche zur Mehrfacherkennung.

4.2 Fiktives Biographiegespräch

Der abschließende Testfall „Fiktives Biographiegespräch“ führt eine vollständige Analyse sowie Genauigkeitsbetrachtung der automati-

sierten Erkennung von Ortsbezügen aus audiovisuellen Inhalten in einem möglichst realitätsnahen Umfeld aus. Zur Analyse wurde ein fiktives, adaptiertes Biographiegespräch mit Fokus auf eine erhöhte Anzahl an Ortsnennungen durchgeführt. Das Video ist in deutscher Sprache aufgezeichnet und beinhaltet insgesamt 96 manuell erhobene Ortsangaben.

Tab. 7 zeigt die Evaluationsergebnisse des Prozessschritts *Transkript* für den lokalen sowie online-basierten Ansatz mit Whisper. Ähnlich wie im vorherigen Testfall zeigen die hier erzielten Ergebnisse, dass die online-basierte Instanz (Trans_{Online}) eine geringere WER (8,3 % vs. 9,1 %) sowie höhere WRR (93,8 % vs. 91,9 %) gegenüber der lokalen Instanz (Trans_{Local}) aufweist. Das bedeutet, dass die Fehleranfälligkeit bei der Verwendung von Whisper_{Online} zum Transkribieren von audiobasierten Inhalten generell geringer ist als bei einer Verwendung von Whisper_{Local}. Gleichzeitig zeigt sich die Wichtigkeit der Einführung des WLER. Im Gegensatz zu den zuvor genannten Parametern sowie zum vorherigen Testfall weist die WLER in der lokalen Whisper-Instanz (93,8 % $\hat{=}$ 90/96) eine um zwei Orte höhere Genauigkeit im Vergleich zur online-basierten Instanz (91,7 % $\hat{=}$ 88/96) auf. Es zeigt sich, dass die Voraussetzungen der für die weiteren Schritte relevanten, korrekt transkribierten Orte mithilfe der lokalen Whisper-Instanz höher sind. In diesem realitätsnahen Testfall werden aufgrund eines geringen Genauigkeitsunterschieds nachfolgend sowohl die lokale als auch die online-basierte Instanz weiter untersucht.

Zur Bestimmung der ortbezogenen Entitäten aus dem zuvor transkribierten Text werden in *Geoannotation* alle Kombinationen gemäß der Erläuterung in Abschnitt 3.2 untersucht. Als Erweiterung zum vorherigen Testfall wird die online-basierte Instanz mit ChatGPT zusätzlich in ihren Parametern angepasst, da sich im Lauf der vorliegenden Untersuchungen vermehrt eine API-seitige Tokenbegrenzung sowie teilweise eine zu explizierte und deterministische Begrenzung der API durch die verwendete *temperature* festgestellt wurde.

Tab. 8 zeigt die Evaluationsergebnisse des Geoannotation für alle aufgezeigten Kombinationsmöglichkeiten. Um auch hier die Genauigkeit der Geoannotation separat zu betrachten, bezieht sich die Berechnung der Gütekriterien auf die korrekt transkribierten Orte ($n = 88$ aus Trans_{Online}; $n = 90$ aus Trans_{Local}) aus dem vorherigen Prozessschritt (vgl. *Tab. 7*). Es zeigt sich, dass generell alle Kombinationen die höheren F1-Scores aufweisen, bei welchen ChatGPT als Geoannotation-Ansatz gewählt wird. Dabei spielt es nur eine zweitrangige Rolle, ob das Transkript via lokaler oder online-basierter Instanz erzeugt wird. Gleichzeitig zeigt sich, dass die Nutzung unterschiedlicher Parameter für *max_tokens* sowie *temperature* (vgl. Online_lowParams vs. Online_highParams) nur sehr geringe Auswirkungen hat. Im Einzelfall werden zwar unterschiedliche Orte korrekt identifiziert, in der Gesamtbetrachtung verändert das den F1-Score jedoch nur geringfügig. Der höchste

	WER (Word Error Rate)		WRR (Word Recognition Rate)		SRR (Sentence Error Rate)		WLER (Word Location Error Rate)	
Trans _{Online}	8,3 %	(152/1837)	93,8 %	(1724/1837)	100 %	(1/1)	91,7 %	(88/96)
Trans _{Local}	9,1 %	(168/1837)	91,9 %	(1688/1837)	100 %	(1/1)	93,8 %	(90/96)

Tab. 7 | Evaluationsergebnisse des Prozessschritts *Transkript* in der Untersuchung „Fiktives Biographiegespräch“

	TP (True Positive)	FP (False Positive)	FN (False Negative)	Precision	Recall	F1-Score
a) Trans _{Local} -Geo _{Online_lowParams}	68	2,5	21,5	0,965	0,760	0,850
*	83	2,5	6,5	0,971	0,927	0,949
b) Trans _{Local} -Geo _{Online_highParams}	69	4	22	0,945	0,758	0,841
*	82	4	9	0,953	0,901	0,927
c) Trans _{Online} -Geo _{Online_lowParams}	64	1,5	26,5	0,977	0,707	0,821
*	79	1,5	11,5	0,981	0,873	0,924
d) Trans _{Online} -Geo _{Online_highParams}	65,5	2,5	25	0,963	0,724	0,826
*	79,5	2,5	11	0,970	0,878	0,922
e) Trans _{Online} -Geo _{Local}	62	1	28	0,984	0,689	0,810
f) Trans _{Local} -Geo _{Local}	61	1,5	27,5	0,976	0,689	0,808

* Mögliches Optimierungspotenzial, wenn ChatGPT mehrfach genannte Orte mehrfach identifizieren würde.

Tab. 8 | Evaluationsergebnisse des Prozessschritts *Geoannotation* in der Untersuchung „Fiktives Biographieggespräch“

F1-Score wird mit der Kombination a) Trans_{Local}-Geo_{Online_lowParams} (F1-Score = 0,850), gefolgt von b) Trans_{Local}-Geo_{Online_highParams} (F1-Score = 0,841) und d) Trans_{Online}-Geo_{Online_highParams} (F1-Score = 0,826) erreicht. Im Gegensatz dazu erreichen die Kombinationen, bei welchen Flair als Geoannotation-Bibliothek verwendet wird, die geringsten F1-Scores. Konkret resultiert aus e) Trans_{Online}-Geo_{Local} ein F1-Score von 0,810 und aus f) Trans_{Local}-Geo_{Local} ein F1-Score von 0,808. Es lässt sich festhalten, dass der Unterschied bei der Anzahl an identifizierten ortsbezogenen Entitäten zwischen den einzelnen Kombinationen und somit zwischen lokaler und online-basierter Instanz nicht so stark variiert (maximale Varianz von 4,2 Prozentpunkten zwischen a) und f)) wie im Testfall „Reisevideo“. In der Ergebnisbetrachtung zeigt sich ebenfalls die große Schwäche in der Erkennung mehrfach genannter Orte via ChatGPT. Die möglichen F1-Scores durch das kalkulierte Optimierungspotenzial (siehe Zeile *) bei den Kombinationen liegen dann konstant über 0,900, was wiederum ein deutlicher Genauigkeitsanstieg gegenüber der lokalen Instanz von Flair ist.

In der genauen Betrachtung der korrekt sowie fehlerhaft identifizierten Orte zeigt sich, dass vor allem kleinere, regional bekannte Orte (z. B. Baiersbronn im Schwarzwald) eine Herausforderung für ChatGPT ist, wohingegen überregional bekannte Orte wie beispielsweise Sehenswürdigkeiten in Hamburg oder Berlin auch durch den Kontextbezug überwiegend korrekt identifiziert werden. Flair hingegen zeigt immer dann eine Schwäche, wenn im fiktiven Biographieggespräch Kontextinformationen fehlen, exemplarisch „*Ich spaziere gem durch den Hafen*“ anstatt „*Ich spaziere gem durch den Hamburger Hafen*“.

Tab. 9 zeigt die Evaluationsergebnisse des Prozessschritts *Geocoding* in den bekannten Kombinationsmöglichkeiten beider Ansätze. Um erneut die Genauigkeit des Geocodings separat zu betrachten, bezieht sich die Berechnung der korrekt geocodierten Orte auf die zuvor identifizierten, ortsbezogenen Entitäten. Die Ergebnisse zeigen im Gegensatz zur vorherigen Analyse, dass der rein lokale Ansatz f) Trans_{Local}-Geo_{Local} die höchste Anzahl korrekt geocodierter Orte (76,2 % $\hat{=}$ 46,5/61) aufweist. Allein d) Trans_{Online}-Geo_{Online_highParams} mit 75,6 % $\hat{=}$ 49,5/65,5 korrekt geocodierter Orte und e) Trans_{Online}-Geo_{Local} mit 71,8 % $\hat{=}$ 44,5/62 korrekt geocodierter Orte weisen gering schwächere Genauigkeits-

werte bei der Geocodierung auf. Besonders die Verwendung von Nominativ ist hier positiv hervorzuheben. In den jeweils anderen kombinierten Ansätzen mit ChatGPT als online-basierte Instanz werden deutlich weniger Orte anteilmäßig korrekt geocodiert. Auch hier wird das mögliche Optimierungspotenzial von ChatGPT berechnet, insofern eine Mehrfacherkennung von genannten Orten stattfinden kann (siehe Zeile *). Im Gegensatz zu den vorherigen Analysen zeigt sich hier, dass lediglich d) Trans_{Online}-Geo_{Online_highParams} die Genauigkeit weiter erhöht, während die Ansätze a), b) und c) nicht die Genauigkeit der lokalen Geocodierungsansätze e) und f) erreichen. In der intensiven Analyse wird deutlich, dass ChatGPT vor allem mit frei erfundenen Koordinaten auffällt. Besonders anzumerken ist, dass eine Anpassung des API-Parameters *temperature* dieses Problem nicht beheben kann. Lediglich die Anpassung des API-Parameters *max_tokens* sorgt für eine höhere Anzahl an korrekt geocodierten Orten.

Abschließend wird auch eine Genauigkeitsuntersuchung des gesamten Workflows angestrebt, nachdem zuvor die einzelnen

	Korrekt geocodierte Orte
a) Trans _{Local} -Geo _{Online_lowParams}	50,0 % (34/68)
*	54,2 % (45/83)
b) Trans _{Local} -Geo _{Online_highParams}	62,3 % (43/69)
*	67,1 % (55/82)
c) Trans _{Online} -Geo _{Online_lowParams}	49,2 % (31,5/64)
*	58,9 % (46,5/79)
d) Trans _{Online} -Geo _{Online_highParams}	75,6 % (49,5/65,5)
*	77,4 % (61,5/79,5)
e) Trans _{Online} -Geo _{Local}	71,8 % (44,5/62)
f) Trans _{Local} -Geo _{Local}	76,2 % (46,5/61)

* Mögliches Optimierungspotenzial, wenn ChatGPT mehrfach genannte Orte mehrfach identifizieren würde.

Tab. 9 | Evaluationsergebnisse des Prozessschritts *Geocoding* in der Untersuchung „Fiktives Biographieggespräch“

	Korrekt identifizierte und geocodierte Orte	
a) Trans _{Local} -Geo _{Online_lowParams}	35,4 %	(34/96)
*	46,9 %	(45/96)
b) Trans _{Local} -Geo _{Online_highParams}	44,8 %	(43/96)
*	57,3 %	(55/96)
c) Trans _{Online} -Geo _{Online_lowParams}	32,8 %	(31,5/96)
*	48,4 %	(46,5/96)
d) Trans _{Online} -Geo _{Online_highParams}	51,6 %	(49,5/96)
*	64,1 %	(61,5/96)
e) Trans _{Online} -Geo _{Local}	46,4 %	(44,5/96)
f) Trans _{Local} -Geo _{Local}	48,4 %	(46,5/96)

* Mögliches Optimierungspotenzial, wenn ChatGPT mehrfach genannte Orte mehrfach identifizieren würde.

Tab. 10 | Evaluationsergebnisse des gesamten Workflows in der Untersuchung „Fiktives Biographiegespräch“

Schritte separat evaluiert und Fehlerquellen identifiziert wurden. Tab. 10 stellt die zusammenfassenden Ergebnisse der korrekt identifizierten und geocodierten Orte dar. Es zeigt sich, dass die höchste Anzahl an Orten mit dem rein online-basierten Ansatz d) Trans_{Online}-Geo_{Online_highParams} (51,6 % $\hat{=}$ 49,5/96) erzielt wird. Der rein lokale Ansatz f) Trans_{Local}-Geo_{Local} weist mit 48,4 % $\hat{=}$ 46,5/96 nur eine leicht geringere Anzahl an korrekt identifizierten und geocodierten Orten auf. Hingegen zeigt sich, dass bei online-basierten Geoparsing-Ansätzen mit geringen API-Parametern (siehe a) und c)) sowie bei der Kombination von lokalem Transkript und online-basiertem Geoparsing (siehe b)) starke Genauigkeitseinbußen zu erwarten sind. Durch Beheben der Schwäche zur Mehrfacherkennung via ChatGPT lässt sich vor allem das Potenzial für Ansatz d) deutlich erhöhen (64,1 % $\hat{=}$ 61,5/96).

5 DISKUSSION

Bei den Untersuchungen der vorgestellten Verfahren sind verschiedene Herausforderungen aufgetreten, aus denen sich Erkenntnisse je Prozessschritt für den praktischen Einsatz ableiten lassen. Um die Nachvollziehbarkeit der gezeigten Untersuchungsergebnisse zu erhöhen, wird eine auftretende Fehlkoordinierung beispielhaft an einer Ortsangabe erläutert.

Im Prozessschritt *Transkript* treten Fehler vermehrt bei der Transkription von Satzzeichen auf. Allerdings zeigen diese in den nachfolgenden Prozessschritten, insbesondere bei der Identifikation von Orten, nur einen geringen Einfluss. Eine während der Untersuchung als effektiv bewertete Vorgehensweise besteht darin, längere audio-basierte Inhalte, beispielsweise an Sprecherpausen oder bei thematischen Veränderungen, aufzuteilen. Dies dient dazu, eine API-seitige Tokenbegrenzung zu verhindern. Des Weiteren zeigt sich eine geringere Fehleranfälligkeit im weiteren Analyseverlauf, wenn der transkribierte Text in Kleinschreibung konvertiert wird. Es besteht zudem eine Fehleranfälligkeit bei der Transkription bestimmter

Buchstabenkombinationen, wie „ss“ vs. „ß“ oder „ie“ vs. „i“, sowie bei unbestimmten und bestimmten Artikeln. Englische Wörter in deutschen Texten weisen eine besonders hohe Fehleranfälligkeit auf, wohingegen Füllwörter wie „ähm“ nicht transkribiert werden. Die Untersuchungen zeigen, dass bei undeutlicher Aussprache und Störgeräuschen in den Audioinhalten die Fehleranfälligkeit des Transkripts steigt. Es lässt sich erkennen, dass sowohl eine lokale als auch online-basierte Whisper-Instanz zum korrekten Transkribieren von Ortsangaben aus biographischen Audioinhalten gewinnbringend eingesetzt werden kann (jeweils > 90 % Genauigkeit). Größere Herausforderungen existieren durch eine undeutliche Aussprache sowie Störgeräusche. Mit Blick auf die Teilnehmergruppe der Biographiegespräche empfiehlt sich daher, ein besonderes Augenmerk auf die Aufnahmequalität sowie Aussprache zu legen, um schlussendlich eine hohe Transkriptgenauigkeit zu gewährleisten. Das unten gezeigte Fehlerbeispiel untermauert diese Aussage.

Der Prozessschritt *Geoannotation* erfordert eine differenzierte Betrachtung der verschiedenen Instanzen. Geo_{Local} (Flair) zeigt beständigere Ergebnisse, da das Modell lokal verfügbar ist. Fehlerhaft transkribierte Orte, insbesondere aufgrund von Buchstabenverwechslungen (z. B. „ie“ vs. „i“), können jedoch nicht immer als ortsbezogene Entitäten identifiziert werden. Darüber hinaus bestätigt sich die Unschärfe zwischen orts- und organisationsbezogenen Entitäten. Demgegenüber präsentiert sich Geo_{Online} (ChatGPT) als Blackbox und listet trotz mehrfacher Nennung von Orten im Text diese jeweils nur einmal als Entität auf. Im Gegensatz zu Geo_{Local} können fehlerhafte transkribierte Orte aufgrund des textlichen Kontexts teilweise dennoch korrekt als ortsbezogene Entität identifiziert werden. Es zeigt sich als genauigkeitssteigernde Maßnahme, Texte vor der API-Anfrage zu teilen, da in der Ausgabe ansonsten Entitäten aufgrund der API-seitigen Tokenbegrenzung nicht identifiziert werden können. Gleichzeitig steigt dadurch das Risiko des Verlusts des textlichen Kontexts, was die Genauigkeit wiederum negativ beeinflussen kann. Die Untersuchungen verdeutlichen, dass eine Erhöhung des Parameters *temperature* bei Geo_{Online} (ChatGPT) zu mehr Varianz in der Ausgabe führt, sodass Orte frei erfunden werden und dadurch die Anzahl der FP steigt. Es zeigt sich insgesamt eine leicht höhere Genauigkeit des rein online-basierten Ansatzes mit ChatGPT_{highParams} im Gegensatz zum lokalen Ansatz mit Flair-Nominativ. Eine Steigerung der Genauigkeit des online-basierten Ansatzes kann künftig durch die Entwicklung eines Ansatzes zur korrekten Identifikation von mehrfach genannten Orten erreicht werden. Insbesondere auf Rechtschreibfehler im Transkript reagiert ChatGPT aufgrund des Kontextwissens mit korrekt identifizierten Orten. Anwender erhalten hier jedoch keinen Einblick in mögliche Zwischenschritte oder Fehler, was gleichzeitig die Gefahr von frei erfundenen Interpretationen, Orten und Koordinaten erhöht. Der lokale Ansatz hingegen zeigt seine Stärke bei der Konsistenz der Ergebnisse, offenbart jedoch Schwächen bei aus dem Transkript resultierenden Rechtschreibfehlern oder nicht exakt benannten Ortsangaben (o. g. Beispiel, (Hamburger) Hafen).

Im Prozessschritt *Geocoding* kristallisieren sich ebenfalls Unterschiede zwischen Geo_{Local} (Nominativ) und Geo_{Online} (ChatGPT) heraus. Geo_{Online} kann textliche Abhängigkeiten zwischen Orten und Aktivitäten herstellen, während Geo_{Local} Entitäten getrennt voneinander koordiniert und somit keine textlichen Abhängigkeiten

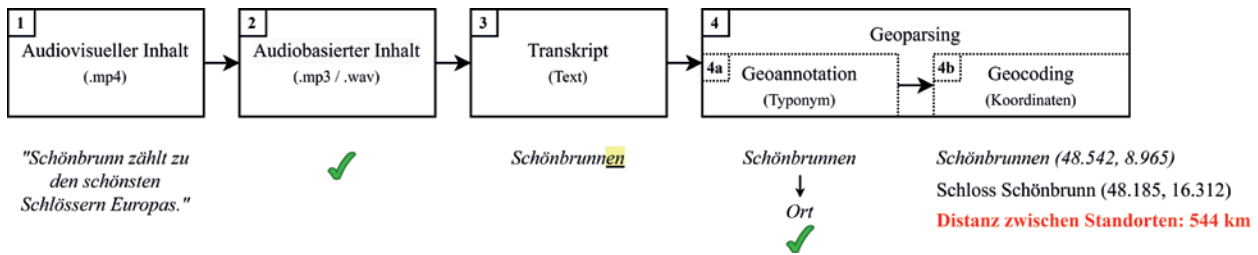


Abb. 2 | Fehlerdarstellung bei der automatisierten Erkennung von Ortsbezügen aus Biographiegesprächen am Beispiel des Schlosses Schönbrunn in Wien

berücksichtigt. Exemplarisch ist der Gesprächsabschnitt „In meiner Freizeit surfte ich gerne auf dem Eisbach.“ zu nennen, in dem ChatGPT den Ortsbezug München zuordnen kann, auch wenn dieser nicht explizit die Benennung München trägt. Allerdings geht diese Stärke damit einher, dass Geo_{Online} dazu neigt, Koordinaten frei zu erfinden, was eine kritische Analyse erfordert. Grundsätzlich zeigen die Untersuchungen daher auch eine deutlich schwächere Koordinatengenauigkeit bei Geo_{Online} im Vergleich zu Geo_{Local}.

Um die Verständlichkeit der dargelegten Forschungsergebnisse zu steigern, soll nachfolgend eine aufgetretene Fehlkoordinierung anhand eines konkreten Ortsbeispiels exemplarisch erläutert werden. Hierzu wird erneut der Testfall „Reisevideo“ betrachtet. Inhalt des Videos ist Wien, sodass in einem Abschnitt das Schloss Schönbrunn mit dem Satz „Schönbrunn zählt zu den schönsten Schlössern Europas.“ benannt wird. Wie Abb. 2 zeigt, entsteht der Fehler durch den Prozessschritt *Transkript*. Durch Hintergrundgeräusche oder eine undeutliche Aussprache wird fälschlicherweise „Schönbrunnen“ anstatt „Schönbrunn“ transkribiert. Im nachfolgenden Schritt der Geoannotation wird jedoch dieser Fehler nicht ermittelt, da auch „Schönbrunnen“ als ortsbezogene Entität identifiziert wird (TP) und folglich eine Geocodierung stattfindet. Erst nach dem letzten Prozessschritt fällt auf, dass das „Schloss Schönbrunn“ aus Wien westlich von Tübingen und somit etwa 544 km entfernt von der eigentlichen Zielkoordinate festgelegt wird. Bei genauer Betrachtung des Orts bei Tübingen lässt sich ein kleiner Bach namens „Schönbrunnen“ identifizieren, welcher irrtümlich geocodiert wurde.

Zusammenfassend zeigt der implementierte Gesamtworkflow zur automatisierten Erkennung von Ortsbezügen aus Biographiegesprächen seine höchste Genauigkeit in der Verwendung des aktuellen rein online-basierten Ansatzes d) Trans_{Online}-Geo_{Online_highParams} (51,6 % der Orte), gefolgt von f) Trans_{Local}-Geo_{Local} (48,4 % der Orte). Wie zuvor erläutert, könnte sich insgesamt die Genauigkeit des online-basierten Ansatzes bei korrekter Identifizierung mehrfach genannter Orte zukünftig deutlich erhöhen. Hier lässt sich der primäre Optimierungsbedarf feststellen, um eine maximale Anzahl an Orten zu erhalten. Gleichzeitig müssen bei der Wahl der Ansätze jedoch auch datenschutzrechtliche Aspekte betrachtet werden. Die Arbeit in Biographiegesprächen adressiert mit Bewohnern in Senioreneinrichtungen eine vulnerable Zielgruppe, bei welcher Gesprächsinhalte zu schützen sind. Je nach Datenschutzrecht kann eine Verwendung eines online-basierten Ansatzes nicht umsetzbar sein, bei welchem die Inhalte in der Cloud analysiert und ausgewertet werden. Die Ergebnisse des implementierten Workflows zeigen, dass die Genauigkeit des rein lokalen Ansatzes im fiktiven Biographiegespräch um nur 3,2 Prozentpunkte schlechter ist als die

maximal erreichte Genauigkeit. Das bedeutet, dass gemäß dieser Untersuchung bei strengen datenschutzrechtlichen Vorgaben die damit enthaltenen Genauigkeitseinbußen vertretbar sind und die Nutzung von f) Trans_{Local}-Geo_{Local} anzustreben ist.

6 ZUSAMMENFASSUNG UND FAZIT

Eine zukünftig steigende Anzahl an Pflegebedürftigen benötigt präventive Maßnahmen zur signifikanten Verbesserung von Lebensqualität und psychischem Wohlbefinden. Dabei gilt die Biographiearbeit durch einen engen, persönlichen Austausch als zielführende Pflegeform. Hierbei können Orte als Anker für Erinnerungen dienen und positive Emotionen auslösen. Die vorliegende Arbeit fokussiert sich auf Ortsbezüge mit direktem Raumbezug in Biographiegesprächen. Dazu wird ein Workflow entwickelt, welcher Ortsbezüge automatisiert erkennt und geocodiert. Verschiedene Testfälle dienen als Basis einer analytischen Untersuchung, in der lokale und online-basierte Ansätze gegenübergestellt und Erkenntnisse für den praktischen Einsatz abgeleitet werden. Dabei stellt sich heraus, dass der online-basierte Ansatz via ChatGPT die höchste Genauigkeit liefert, jedoch Probleme bei Mehrfachnennung von Orten hat oder Koordinaten fiktiv erfindet. Der lokale Ansatz via Flair-Nominativ weist im Vergleich dazu zwar eine leicht geringere Genauigkeit auf, besitzt dafür aber seine Stärke in der Erkennungsbeständigkeit und der Berücksichtigung des Datenschutzes. Bei der im Anwendungsfall konfrontierten, vulnerablen Gesprächsgruppe gilt daher eine individuelle Abwägung der Genauigkeitsanforderungen gegenüber dem Schutz persönlicher Informationen. Abgeleitete Erkenntnisse identifizieren Optimierungspotenziale im Workflow.

In zukünftigen Arbeiten wird der geschaffene Workflow in realen Biographiegesprächen mit Bewohnern in Senioreneinrichtungen eingesetzt, um eine Übertragbarkeit ins Pflegeumfeld zu gewährleisten. Voraussetzung dafür ist die erfolgreiche Prüfung der Ethikkommission, welche zum Zeitpunkt der hier vorliegenden Untersuchungen nicht abgeschlossen ist. Aus technischer Sicht betrachtet empfiehlt es sich zudem, die angesprochenen Probleme der Erkennung mehrfach genannter Orte mit ChatGPT zu adressieren. Gleichzeitig zeigt die durchgeführte Untersuchung den Bedarf an weiteren Forschungsarbeiten, welche Ortsidentifikationen bei NER-Modellen wie Flair auch durch textliches Kontextverständnis gewährleistet. Darüber hinaus wird für die Biographiearbeit angestrebt, die Gesprächsinhalte auch hinsichtlich Emotionen mittels Sentimentanalysen auszuwerten. Emotionen können sowohl aus dem initialen audiovisuellen Eingangsmedium als auch aus dem Transkript abgeleitet werden. Derselbe Ort kann durch eine unterschiedliche Gestik,

Mimik oder Emotion einen komplett anderen persönlichen Bezug erhalten. Langfristiges Ziel sollte daher sein, die identifizierten Emotionen mit den erkannten Ortsangaben zu kombinieren und nutzerspezifisch darzustellen. Die Kommunikation der raumbezogenen und emotionalen Untersuchungsergebnisse stellt eine große Herausforderung dar. Dazu bieten geovisuelle Techniken, wie beispielsweise nutzergeleitete Leitung mit Storytelling /Rolwes et al. 2023/, vielversprechende Ansätze, welche die Grenzen der kognitiven Belastbarkeit der Senioren berücksichtigen.

DANK

Dieser Forschungsbeitrag wird gefördert durch die Bund-Länder-Initiative „Innovative Hochschule“ und ist Teil des Verbundprojekts „EMPOWER“ von fünf Hochschulen für Angewandte Wissenschaften in Rheinland-Pfalz.

LITERATUR

- Alharbi, S.; Alrazgan, M.; Alrashed, A.; Alnomasi, T.; Almojel, R.; Alharbi, R. et al. (2021): Automatic Speech Recognition: Systematic Literature Review. In: IEEE Access 9(2021), 131858 – 131876. DOI: 10.1109/ACCESS.2021.3112535.
- Arora, S. J.; Singh, R. P. (2012): Automatic Speech Recognition: A Review. In: International Journal of Computer Applications 60(2012)9.
- Bauer, J. J.; McAdams, D. P.; Pals, J. L. (2008): Narrative identity and eudaimonic well-being. In: Journal of Happiness Studies 9(2008)1, 81 – 104. DOI: 10.1007/s10902-006-9021-6.
- Benzeghiba, M.; Mori, R. de; Deroo, O.; Dupont, S.; Erbes, T.; Jouvét, D. et al. (2007): Automatic speech recognition and speech variability: A review. In: Speech Communication 49(2007)10-11, 763 – 786. DOI: 10.1016/j.specom.2007.02.006.
- Böhm, K.; Tesch-Römer, C.; Ziese, T. (Hrsg.) (2009): Gesundheit und Krankheit im Alter. Beiträge zur Gesundheitsberichterstattung des Bundes. Robert-Koch-Institut, Berlin. https://www.statistischebibliothek.de/mir/receive/DEMonografie_mods_00000117 (13. 11. 2023).
- Corsten, S.; Lauer, N. (2020): Biography work in in long-term residential aged care with tablet support to improve the quality of life and communication – study protocol for app development and evaluation / Biografiearbeit in Senioreneinrichtungen mit Tablet-Unterstützung zur Verbesserung der Lebensqualität und Kommunikation (BaSeTaLK) – Studienprotokoll zur App-Entwicklung und Evaluation. In: International Journal of Health Professions, 7(2020)1, 13 – 23. DOI: 10.2478/ijhp-2020-0002.
- Cowie, J.; Lehnert, W. (1996): Information extraction. In: Communications of the ACM 39(1996)1, 80 – 91. DOI: 10.1145/234173.234209.
- Drageset, J.; Kirkevold, M.; Espehaug, B. (2011): Loneliness and social support among nursing home residents without cognitive impairment: a questionnaire survey. In: International Journal of Nursing Studies 48(2011)5, 611 – 619. DOI: 10.1016/j.ijnurstu.2010.09.008.
- Fraher, A.; Coffey, A. (2011): Older people's experiences of relocation to long-term care. In: Nursing Older People 23(2011)10, 23 – 27. DOI: 10.7748/nop2011.12.23.10.23.c8838.
- Freudenstein, U.; Jagger, C.; Arthur, A.; Donner-Banzhoff, N. (2001): Treatments for late life depression in primary care – a systematic review. In: Family Practice 18(2001)3, 321 – 327. DOI: 10.1093/fampra/18.3.321.
- Gennip, D. van; Hoven, E. van den; Markopoulos, P. (2015): Things That Make Us Reminisce. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 18.–23.04.2015. Seoul, Republic of Korea. ACM, New York, NY, 3443 – 3452.
- Giordano, K.; Lauer, N.; Leusch, V.; Kreiter, D.; Corsten, S. (2023): Tablet-based biography work in long-term care homes to support quality of life and prevent depression. In: Educational Gerontology (2023), 1 – 14. DOI: 10.1080/03601277.2023.2225362.
- Gregory, I.; Donaldson, C.; Murrieta-Flores, P.; Rayson, P. (2015): Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research. In: International Journal of Humanities and Arts Computing 9(2015)1, 1 – 14. DOI: 10.3366/ijhac.2015.0135.
- Gritta, M.; Pilehvar, M. T.; Collier, N. (2020): A pragmatic guide to geoparsing evaluation: Toponyms, Named Entity Recognition and pragmatics. In: Language Resources and Evaluation 54(2020)3, 683 – 712. DOI: 10.1007/s10579-019-09475-3.
- Hu, Y.; Mai, G.; Cundy, C.; Choi, K.; Lao, N.; Liu, W.; Lakhpanal, G.; Zhou, R. Z.; Joseph, K. (2023): Geo-knowledge-guided GPT models improve the extraction of location descriptions from disaster-related social media messages. In: International Journal of Geographical Information Science 37(2023)11, 2289 – 2318. DOI: 10.1080/13658816.2023.2266495.
- Hu, X.; Zhou, Z.; Li, H.; Hu, Y.; Gu, F.; Kersten, J.; Fan, H.; Klan, F. (2022): Location reference recognition from texts: A survey and comparison. arXiv:2207.01683v1. <https://arxiv.org/pdf/2207.01683.pdf> (13. 11. 2023).
- Hunt, M. J. (1990): Figures of merit for assessing connected-word recognisers. In: Speech Communication 9(1990)4, 329 – 336.
- ISO 19111:2019: Geographic information – Referencing by coordinates (ISO 19111:2019) (English version EN ISO 19111:2020). Genf.
- Kanjo, E.; Younis, E. M.; Ang, C. S. (2019): Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. In: Information Fusion 49(2019), 46 – 56. DOI: 10.1016/j.inffus.2018.09.001.
- Karimzadeh, M.; Pezanowski, S.; MacEachren, A. M.; Wallgrün, J. O. (2019): GeoTxt: A scalable geoparsing system for unstructured text geolocation. In: Transactions in GIS 23(2019)1, 118 – 136. DOI: 10.1111/tgis.12510.
- Klakow, D.; Peters, J. (2002): Testing the correlation of word error rate and perplexity. In: Speech Communication 38(2002)1-2, 19 – 28.
- Kurzweil, R. (1999): The Age of Intelligent Machines. 3. Aufl. MIT Press, Cambridge, Mass.
- Luanagh, C. Ó.; Lawlor, B. A. (2008): Loneliness and the health of older people. In: International Journal of Geriatric Psychiatry 23(2008)12, 1213 – 1221. DOI: 10.1002/gps.2054.
- Menn, L.; Corsten, S.; Lauer, N.; Wallace, S. J. (2020): The Effectiveness of Biographical Approaches in Long-Term Care: A Systematic Review. In: The Gerontologist 60(2020)4, e309 – e328. DOI: 10.1093/geront/gnz074.
- Metoyer, R.; Zhi, Q.; Janczuk, B.; Scheirer, W. (2018): Coupling Story to Visualization. In: Berkovsky, S.; Hijikata, Y.; Rekimoto, J.; Burnett, M.; Billinghurst, M.; Quigley, A. (Hrsg.): 23rd International Conference on Intelligent User Interfaces, 07.–11.03.2018, Tokyo, Japan. ACM, New York, NY, 503 – 507.
- Nadeau, D.; Sekine, S. (2007): Named Entities: Recognition, classification and use. In: Lingvisticae Investigationes 30(2007)1, 3 – 26. DOI: 10.1075/li.30.1.03nad.
- Nominatim developer community (o. J.): Nominatim Manual. <https://nominatim.org/release-docs/develop/api/Overview/> (11. 11. 2023).
- OpenAI (2023): GPT-4 Technical Report. <https://arxiv.org/pdf/2303.08774.pdf> (13. 11. 2023).
- OpenAI Platform (2023): Documentation – GPT models. <https://platform.openai.com/docs/guides/gpt/managing-tokens> (30. 10. 2023).
- Porwal, K.; Srivastava, H.; Gupta, R.; Pratap Mall, S.; Gupta, N. (2022): Video Transcription and Summarization using NLP. In: Proceedings of the Advancement in Electronics & Communication Engineering 2022. Social Science Research Network. DOI: 10.2139/ssrn.4157647.

Potamianos, G.; Neti, C.; Luettin, J.; Matthews, I. (2004): Audio-Visual Automatic Speech Recognition: An Overview. In: Issues in Visual and Audio-Visual Speech Processing. MIT Press, Cambridge, Mass.

Purves, R. S.; Clough, P.; Jones, C. B.; Hall, M. H.; Murdock, V. (2018): Geographic Information Retrieval: Progress and Challenges in Spatial Search of Text. In: FNT in Information Retrieval 12(2018)2-3, 164–318. DOI: 10.1561/15000000034.

Rabiner, L. R. (1989): A tutorial on hidden Markov models and selected applications in speech recognition. In: Proceedings of the IEEE 77(1989)2, 257–286. DOI: 10.1109/5.18626.

Radford, A.; Kim, J. W.; Xu, T.; Brockman, G.; McLeavey, C.; Sutskever, I. (2022): Robust Speech Recognition via Large-Scale Weak Supervision. arXiv:2212.04356v1. <https://arxiv.org/pdf/2212.04356.pdf> (13. 11. 2023).

Rolwes, A.; Stockemer, J.; Böhm, K. (2023): Analyzing Spatio-Temporal Correlations with User-Oriented Guidance – An Interactive Visualization Approach for Demand-Oriented Limited Service Offers. In: 27th International Conference Information Visualisation, 24.–28.07.2023, Tampere, Finnland. DOI: 10.1109/IV60283.2023.00046.

Ruthmann, S.; Rolwes, A.; Böhm, K. (2020): Räumliche Verortung von textbasierten Social-Media-Einträgen am Beispiel von Polizei-Tweets. In: FOSSGIS e. V. (Hrsg.): FOSSGIS Tagungsband 2020. Freiburg, 83–86.

Serere, H. N.; Kanilmaz, U. N.; Ketineni, S.; Resch, B. (2023): A Comparative Study of Geocoder Performance on Unstructured Tweet Locations. In: GI_Forum 2023 11(2023)1, 110–117. DOI: 10.1553/giscience2023_01_s110.

Sesselmann, M.; Naber, T.; Stricker, R.; Scheller, S. (2023): Mobiles Laser-scanning und Deep Learning zur automatischen Erfassung des Oberflächenzustands kommunaler Straßen. allgemeine vermessungs-nachrichten 130(2023)1-2, 21–33.

Specht-Tomann, M. (2018): Biografiearbeit: in der Gesundheits-, Kranken- und Altenpflege. 3. Aufl. Springer, Berlin.

Statista (2023): Anzahl Pflegebedürftige in Deutschland bis 2021 (veröffentlicht von Rainer Radtke). <https://de.statista.com/statistik/daten/studie/2722/umfrage/pflegebeduerftige-in-deutschland-seit-1999/> (05.09.2023).

theTravellers (2016): Wien in 5 Minuten | Reiseführer | Die besten Sehenswürdigkeiten. https://www.youtube.com/watch?v=_cJJoGh3Hqo (02.11.2023).

Yu, D.; Deng, L. (2015): Automatic speech recognition. A deep learning approach. Springer, London.

AUTOREN



M. Sc. Alexander Rolwes

HOCHSCHULE MAINZ
I3MAINZ, INSTITUT FÜR RAUMBEZOGENE
INFORMATIONEN- UND MESSTECHNIK

Lucy-Hillebrand-Str. 2 | 55128 Mainz | Deutschland
alexander.rolwes@hs-mainz.de | ORCID: 0000-0001-9599-0219



M. Eng. Patrick Stellbauer

HOCHSCHULE MAINZ
FACHBEREICH TECHNIK,
FACHRICHTUNG ANGEWANDTE
INFORMATIK UND GEODÄSIE

Lucy-Hillebrand-Str. 2 | 55128 Mainz | Deutschland
patrick.stellbauer@hs-mainz.de | ORCID: 0009-0002-2622-0944



M.A. Almut Plath

KATHOLISCHE HOCHSCHULE MAINZ

Saarstr. 3 | 55122 Mainz | Deutschland
almut.plath@kh-mz.de | ORCID: 0000-0002-2795-4301



Prof. Dr.-Ing. Klaus Böhm

HOCHSCHULE MAINZ
I3MAINZ, INSTITUT FÜR RAUMBEZOGENE
INFORMATIONEN- UND MESSTECHNIK

Lucy-Hillebrand-Str. 2 | 55128 Mainz | Deutschland
klaus.boehm@hs-mainz.de | ORCID: 0009-0008-8492-7650



Prof. Dr.-Ing. Markus Schaffert

HOCHSCHULE MAINZ
I3MAINZ, INSTITUT FÜR RAUMBEZOGENE
INFORMATIONEN- UND MESSTECHNIK

Lucy-Hillebrand-Str. 2 | 55128 Mainz | Deutschland
markus.schaffert@hs-mainz.de | ORCID: 0000-0002-7970-9164



Prof. i.K. Dr. rer. medic. Sabine Corsten

KATHOLISCHE HOCHSCHULE MAINZ

Saarstr. 3 | 55122 Mainz | Deutschland
sabine.corsten@kh-mz.de | ORCID: 0000-0003-3695-8010



Prof. Dr. Pascal Neis

HOCHSCHULE MAINZ
FACHBEREICH TECHNIK,
FACHRICHTUNG ANGEWANDTE
INFORMATIK UND GEODÄSIE

Lucy-Hillebrand-Str. 2 | 55128 Mainz | Deutschland
pascal.neis@hs-mainz.de | ORCID: 0000-0002-5158-796X

Manuskript eingereicht: 9. 11. 2023 | Im Peer-Review-Verfahren begutachtet