

Wie Sprach-KI die Barrierefreiheit unterstützt

MT für Leichte Sprache

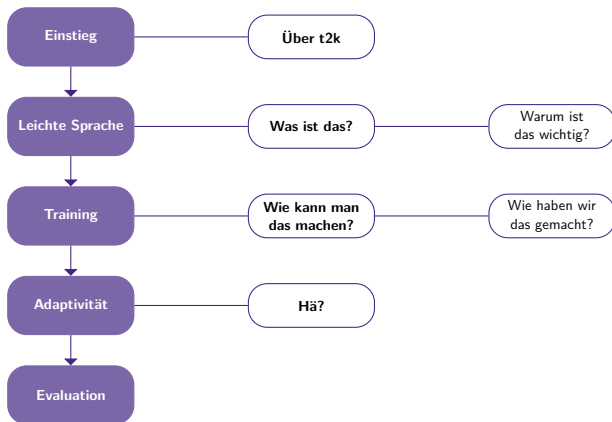
Dr. Anne-Kathrin Schumann



27.8.2024



Plan für heute.



t2k GmbH: Kurzsteckbrief

Spezialist für KI-basierte Sprachverarbeitung aus Dresden



Gründung

Gründung 2021 nach mehrjähriger Forschungstätigkeit mit Unterstützung des Saxeed-Netzwerks Freiberg, gefördert durch das BMWK und die SAB



Auszeichnungen

Nominiert für den Sächsischen Gründerpreis, Preisträger des Civic Innovation Wettbewerbs des BMAS



Wachstum

Stabiles, stetiges Wachstum, profitabel seit Tag 1. Mittlerweile ein Team von 9 Expert:innen

t2k GmbH: Kurzsteckbrief

Wissen automatisieren mit Künstlicher Intelligenz und automatischer Sprachverarbeitung



Dokumenten-automatisierung

Dokumenten-Workflows mit generativer KI, Bild-zu-Text (OCR), Sprache-zu-Text (Voice)



Inhaltliche und sprachliche Textoptimierung mit NLP

Texte in barrierefreie, leicht lesbare Sprache übersetzen



Individuelle NLP-Entwicklung

Beratung und Umsetzung individueller NLP-Anwendungsfälle

Alleinstellungsmerkmale

Wissensgetriebene Organisation

- Nachweisbare einschlägige Expertise: 3 promovierte Computerlinguisten im Team, zusätzlich Experten für Machine Learning und Computerlinguistik
 - ▶ Zusammen mehr als 30 Jahre einschlägige Erfahrung allein im NLP-Team
 - ▶ Mehr als 1000 Zitate auf eigenen NLP-Forschungspublikationen (Google Scholar)
- Methodenvielfalt auch im Low-Resource-Bereich
- Tiefes Verständnis der linguistischen Hintergründe



Inhouse-NLP-Entwicklung

- Automatische Sprachvereinfachung
- Entitätenerkennung und Anonymisierung: IBAN, Steuernummer, HRN, Straße, ...
- Textzusammenfassung
- Neuronale Suche
- RAG
- Agentenbasierte Systeme



Leichte Sprache: Was ist das?

Vision

Leichte Sprache automatisch für alle und überall

Texteingabe

Mit der Wohnberatungsstelle im Landratsamt unterstützt der Landkreis Wunsiedel i. Fichtelgebirge die Bürgerinnen und Bürger dabei, möglichst lange in der vertrauten Umgebung wohnen zu bleiben.

Leichte Sprache

Einfache Sprache

Leichte Sprache: Was ist das?

Vision

Leichte Sprache automatisch für alle und überall

Texteingabe

Mit der Wohnberatungsstelle im Landratsamt unterstützt der Landkreis Wunsiedel i. Fichtelgebirge die Bürgerinnen und Bürger dabei möglichst lange in der vertrauten Umgebung wohnen zu bleiben.

Ausgabe

Die Wohnberatungsstelle im Kreis-Amt Wunsiedel hilft den Menschen:

- Sie sollen möglichst lange in der gewohnten Umgebung wohnen bleiben.

Leichte Sprache

Einfache Sprache

Leichte Sprache: Was ist das?

t2k

Einstellungen

Abmelden

Texteingabe

Leichte Sprache →

Einfache Sprache →

Der mit der Konvention verbundene Paradigmenwechsel weg von Fürsorge und Integration hin zur Inklusion betrifft alle Menschen, egal ob mit oder ohne Behinderungen, und nahezu jeden Lebensbereich: von der Infrastruktur über die Bildung und den Arbeitsmarkt bis hin zur Gestaltung der Freizeitangebote – überall wird an der Realisierung der Inklusion gearbeitet.

Ausgabe



Das bedeutet: Alle Menschen mit und ohne Behinderung sollen sich besser um die Menschen mit Behinderung kümmern. Und das in fast jedem Lebens-Bereich. Zum Beispiel:

- In der Schule
- in der Ausbildung,
- im Arbeits-Markt,
- in der Freizeit.



Warum ist Verständlichkeit wichtig?

**In Deutschland leben 3 Millionen Legastheniker.
12% der erwerbsfähigen Bevölkerung können nur
unzureichend schreiben und lesen.**



Lesen ist grundlegend für gesellschaftliche Teilhabe und Integration.



Das Behindertengleichstellungsgesetz fordert die Nutzung der Leichten Sprache.



Die manuelle Übersetzung in Leichtes Deutsch ist teuer und zeitaufwendig.

Wie kann man das automatisieren? Variante 1: Prompting

- Eingabe: „Mehr als 90 Prozent der älteren Menschen leben in ihren vertrauten Wohnungen oder Häusern und möchten dort auch bleiben. Aber die meisten Wohnungen und Häuser entsprechen den veränderten Bedürfnissen des Alters nicht. Der Einstieg in die Badewanne ist zu hoch, die Stufen am Eingangsbereich können nicht mehr überwunden werden oder der Balkon wird aufgrund der Schwelle nicht mehr genutzt. Mit Hilfe von Wohnungsanpassungsmaßnahmen können diese Defizite ausgeglichen werden.“

Wie kann man das automatisieren? Variante 1: Prompting

- Eingabe: „Mehr als 90 Prozent der älteren Menschen leben in ihren vertrauten Wohnungen oder Häusern und möchten dort auch bleiben. Aber die meisten Wohnungen und Häuser entsprechen den veränderten Bedürfnissen des Alters nicht. Der Einstieg in die Badewanne ist zu hoch, die Stufen am Eingangsbereich können nicht mehr überwunden werden oder der Balkon wird aufgrund der Schwelle nicht mehr genutzt. Mit Hilfe von Wohnungsanpassungsmaßnahmen können diese Defizite ausgeglichen werden.“
- ChatGPT4 - mit den Regeln der DinSpec: "Viele ältere Menschen leben in ihren eigenen Wohnungen oder Häusern. Sie möchten dort auch im Alter bleiben. Aber oft passen die Wohnungen nicht zu den neuen Bedürfnissen. Zum Beispiel ist die Badewanne zu hoch. Oder es gibt Stufen, die schwer zu gehen sind. Man kann die Wohnung so verändern, dass sie besser passt.“

LS durch Prompting - Probleme

- Entspricht nicht den Regeln der LS (z. B. Kompositatrennung, Textgliederung, Schlüsselwörter)
- Halluzinationen, inhaltlich nicht präzise
- Teilweise großer Informationsverlust oder Verlust des roten Fadens und Überlänge
- Fachliche Anpassbarkeit nicht gegeben
- Abhängigkeit von einem externen Service (Latenz, Skalierbarkeit der Kosten, Verfügbarkeit)

<https://eamt2024.github.io/proceedings/vol1.pdf> (S. 469 ff.)

Eine alte Übersetzung von uns

„Mehr als 90 Prozent der älteren Menschen wohnen in ihren vertrauten Häusern.

Aber die meisten Wohnungen sind nicht für das Alter gemacht.

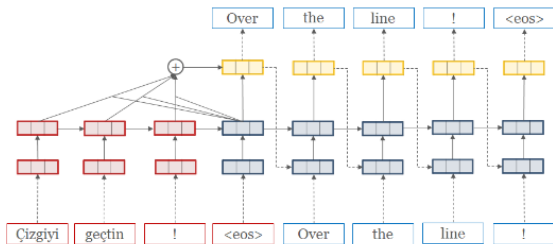
Das bedeutet:

Der Einstieg in die Badewanne ist zu hoch. Oder der Balkon kann nicht mehr benutzt werden!

Das kann man mit Hilfe von Wohnraum-Anpassungen ausgleichen.“

- Klare, regel-konforme Textgliederung
- Schlüsselwörter
- Informationsreduktion
- Inhaltliche Präzision, keine Halluzinationen
- Trennung zusammengesetzter Wörter (eigenes Modell)
- Erkennung und Ersetzung verbleibender schwerer Wörter (eigene Modelle)

Satzparalleles Training - neuronale maschinelle Übersetzung

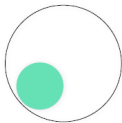


1. Texte werden satzweise zerlegt.
2. Ausgangs- und Zielsätze werden gemeinsam in das Modell gegeben (Training).
3. Das Modell lernt, einen inhaltsgleichen Satz in der Zielsprache zu erzeugen (Inferenz).

„Schwesterprojekt“: MT Sorbisch-Deutsch-Sorbisch für das Witaj-Sprachzentrum Bautzen (Sotra)

(Quelle des Bildes: <https://opennmt.net/>)

Nachhaltigkeit & Sicherheit



Daten- und Ressourceneffizienz

Die t2k-Modelle sind im Vergleich mindestens 100x kleiner als bekannte Modelle von OpenAI (GPT4 oder ChatGPT).

Dadurch ist das Training unserer spezialisierten Modelle deutlich weniger ressourcen- und kostenintensiv als das Training von LLMs.



Inhaltsgleichheit

Durch unseren eigenen technischen Ansatz konditionieren wir unser Modell auf Inhaltsgleichheit.

Unser spezialisierter Ansatz reduziert Falschinformationen und Halluzinationen.

Eigene Modelle – made in Saxony, kombiniert mit wissenschaftlichem Anspruch



Transparenz und Erklärbarkeit der KI

Eigene Entwicklung der Modelle sowie eigene Daten, Benchmarking, wissenschaftliche Veröffentlichungen



Konformität mit aktuellen Gesetzen: DSGVO- sowie EU-AI-Act-konform

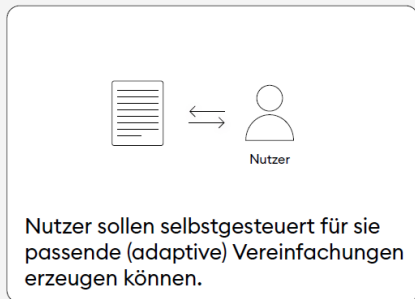
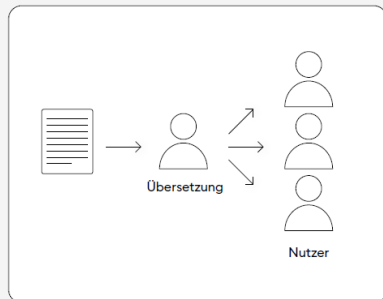
Eine Auditierbarkeit der t2k-Modelle ist möglich und vorgesehen



Lokale Einsetzbarkeit

Modelle können lokal laufen (PC oder Rechenzentrum). Dies ermöglicht den Umgang mit Dokumenten die der Geheimhaltung unterliegen

Perspektivwechsel: Von der Bereitstellung von Verständlichkeit durch Dritte hin zur selbstgesteuerten Textvereinfachung.



Adaptivität - ein Spiel

Dieser Aktionsplan beinhaltet sowohl die bereits umgesetzten Maßnahmen des Landes zugunsten von Menschen mit Behinderungen als auch Maßnahmen, die noch zur Umsetzung vorgesehen sind.



Können wir die Übersetzungen aufsteigend nach Schwierigkeit ordnen?

- Im Landes-Plan stehen zum Beispiel:
 - Was das Land schon für Menschen mit Behinderungen gemacht hat.
 - Und was noch passieren soll.



Können wir die Übersetzungen aufsteigend nach Schwierigkeit ordnen?

- Im Landes-Plan stehen zum Beispiel: ■ Was das Land schon für Menschen mit Behinderungen gemacht hat. ■ Und was noch passieren soll.
- In dem Aktionsplan stehen die schon gemachten Maßnahmen des Landes für Menschen mit Behinderungen. Und die Maßnahmen, die noch kommen sollen.



Können wir die Übersetzungen aufsteigend nach Schwierigkeit ordnen?

- Im Landes-Plan stehen zum Beispiel: ■ Was das Land schon für Menschen mit Behinderungen gemacht hat. ■ Und was noch passieren soll.
- In dem Aktionsplan stehen die schon gemachten Maßnahmen des Landes für Menschen mit Behinderungen. Und die Maßnahmen, die noch kommen sollen.
- In diesem Aktionsplan stehen die bereits umgesetzten Maßnahmen des Landes für Menschen mit Behinderungen.



Können wir die Übersetzungen aufsteigend nach Schwierigkeit ordnen?

- Im Landes-Plan stehen zum Beispiel: ■ Was das Land schon für Menschen mit Behinderungen gemacht hat. ■ Und was noch passieren soll.
- In dem Aktionsplan stehen die schon gemachten Maßnahmen des Landes für Menschen mit Behinderungen. Und die Maßnahmen, die noch kommen sollen.
- In diesem Aktionsplan stehen die bereits umgesetzten Maßnahmen des Landes für Menschen mit Behinderungen.
- Im Aktionsplan stehen zum Beispiel: ■ Was das Land schon für Menschen mit Behinderungen getan hat. ■ Oder was noch passieren soll.



Wie und was evaluieren wir?

- Klassische MT- und Lesbarkeitsmetriken (Bleu, Flesch) haben wir nach Korrelationsstudien verworfen.
- Inhaltliche Korrektheit: Trainierte Metriken (Klassifikation)
- Einfachheit: Eigenes Modell wird entwickelt
- Nutzerstudie: <https://forms.office.com/e/2tu1pvAkBw>,
<https://forms.office.com/e/RR96ZTNQhs>

Ein Papiermacher ist ein Handwerker , der Papier herstellt . In der Gegenwart ist er in einer Papiermühle mit entsprechenden Produktionseinrichtungen (Industrielle Papierfabrik) tätig . Seit dem Jahr 2005 heißt der Beruf nach der Klassifikation in Deutschland Papiertechnologe . In der größten Zahl der Fälle hat jeder leitende Papiermüller ein Wasserzeichen verwendet , das allein für seine Wirkungszeit typisch war .

Ein Papier-Macher macht Papier . Heute arbeitet er in einer Papier -- Mühle . Seit dem Jahr 2005 heißt der Beruf Papier-Technologe . In den meisten Fällen hat jeder leitende Papier-Müller ein Wasser-Zeichen benutzt .



Unser Team

Dr. Jesús Calvillo, Felix Dittrich, Dennis Engel, Anna-Janina Goecke, Ian List, Dr. Umesh Patil, Martin Rosenbusch, Dr. Anne-Kathrin Schumann, Tobias Wittig

<https://portal.text2knowledge.de>

info@text2knowledge.de

