

TXT-Sprache als Spur

Entwicklung eines KI-basierten Tools zur Bedarfsanalyse forensisch-linguistischer Expertise

Programm / Ausschreibung	KIRAS, Kooperative F&E-Projekte, KIRAS-Kybernet-Pass CS Kooperative F&E Projekte (CS KFE_2024)	Status	laufend
Projektstart	01.01.2026	Projektende	31.12.2027
Zeitraum	2026 - 2027	Projektlaufzeit	24 Monate
Projektförderung	€ 683.648		
Keywords	Forensische Linguistik; Kriminaltechnik; Künstliche Intelligenz; Textanalysen; Autorenerkennung; Profilerstellung		

Projektbeschreibung

Die forensische Sprachwissenschaft ist eine sich im Aufschwung befindliche Disziplin, die dann zur Anwendung kommt, wenn Schriftstücke jeglicher Art zum Gegenstand von Ermittlungen werden und betrifft unter anderem strafbares Verhalten wie Stalking, Erpressung, Hasspostings und üble Nachrede. Zusätzlich können auch anonyme Hinweise, Bekennerschreiben und Manifeste Gegenstand von Analysen sein. Wenn, wie oft in solchen Fällen, die einzige Spur zum Täter eine sprachliche Spur ist, kann die forensische Sprachwissenschaft sprachliche Merkmale unbekannter Autor:innen feststellen, diese mit Vergleichstexten potentiell Verdächtiger abgleichen, und Sprachprofile für weitere Ermittlungen erstellen. Da diese Analysen manuell von Expert:innen durchgeführt werden müssen, kommen solche Analysen aktuell nur sehr eingeschränkt zum Einsatz. Es ist daher besonders wichtig zu wissen, in welchen Situationen Expert:innen zu Rate gezogen werden sollen – und genau an diesem Punkt setzt das vorliegende Projekt an.

Das Projekt „TXT – Sprache als Spur“ soll ein KI-gestütztes Analysetool schaffen, das Texte auf ihre Eignung hin untersucht, ob tieferegehende Analysen von Expert:innen möglich und sinnvoll sind. Ergänzend soll eine Voreinschätzung abgegeben werden, ob bereits eine ähnliche sprachliche Spur in der Datenbank vorhanden ist, die dann einer weiteren Analyse durch Expert:innen unterzogen werden könnte. Als Datengrundlage für dieses Projekt soll die handschriftliche Lebenslaufsammlung des Referats Urkunden- & Handschriftuntersuchung im Büro für Kriminaltechnik des Bundeskriminalamts (BK) im Bundesministerium für Inneres (BMI) digitalisiert werden und durch weitere inkriminierte Schreiben ergänzt werden. Methodisch soll dieses Ziel durch den Einsatz von Techniken des maschinellen Lernens erreicht werden, wodurch ein KI-basiertes Tool in diesen beiden Bereichen zeiteffizient eine Aussage darüber treffen kann, ob eine Analyse möglich ist, und ob interne Vergleichstexte näher untersucht werden sollten.

Abstract

Forensic linguistics is an emerging discipline that is used when documents of any kind become the subject of an investigation and thus concerns criminal offenses such as stalking, blackmail, hate postings and defamation. In addition,

anonymous tips, letters of confession and manifestos can also be the focus of analyses. If, as is often the case in such investigations, the only lead to the perpetrator is a linguistic one, forensic linguistics can determine characteristics of unknown authors, compare them to other texts of possible suspects and create language profiles for further investigation. As these analyses have to be carried out manually by experts, their current use is limited. It is therefore particularly important to know in which situations experts should be consulted - and this is precisely where this project comes in.

The "TXT - Language as a Trace" project aims to create an AI-supported analysis tool that examines texts to determine whether in-depth analyses by experts are possible and worthwhile. In addition, a preliminary assessment will be made whether a similar linguistic trace already exists in the database, which could then be subjected to further analysis by experts. As a data basis for this project, the handwritten CV collection of the Document & Handwriting Investigation Unit in the Department of Forensic Science of the Criminal Intelligence Service Austria (BK) in the Federal Ministry of the Interior (BMI) will be digitized and supplemented by further incriminated texts. Methodologically, this goal is to be achieved through the use of machine learning techniques, whereby an AI-based tool can make a time-efficient statement in these two areas whether an analysis is possible and whether comparison texts should be analyzed in greater depth.

Projektkoordinator

- Universität Graz

Projektpartner

- WareTec IT Solutions GmbH
- Bundesministerium für Inneres
- Technische Universität Wien