

# K.Rex

Knowledge Recognition for Evidence eXtraction

<b>Programm / Ausschreibung</b>	KIRAS, Kooperative F&E-Projekte, KIRAS Kooperative F&E-Projekte 2018	<b>Status</b>	laufend
<b>Projektstart</b>	01.12.2019	<b>Projektende</b>	30.11.2021
<b>Zeitraum</b>	2019 - 2021	<b>Projektlaufzeit</b>	24 Monate
<b>Keywords</b>	Multimodale Inhalts-Erschließung; Deep Learning; effiziente Annotation; zweidimensionales Word Embedding; semantische Dokumentsegmentierung		

## Projektbeschreibung

In Ermittlungsfällen der Strafverfolgung und Kriminalitätsbekämpfung fallen große Mengen an heterogenen Dokumenten (z.B. aus Beschlagnahmungen, Konto-Öffnungen) an, die von Ermittlerinnen und Ermittlern gesichtet und hinsichtlich ihrer inhaltlichen (strafrechtlichen) Relevanz bewertet werden müssen. Ohne maschinelle Unterstützung ist das stark steigende Datenaufkommen nicht zu bewältigen. Während der Mensch ganz selbstverständlich Schlussfolgerungen aus Struktur, optischer Anmutung und bildlichen Elementen wie Logos, Stempel oder handschriftlichen Ergänzungen in Dokumenten zieht, bleibt diese Information textbasierten maschinellen Verfahren verborgen. Ist der Text fehlerbehaftet (etwa aus OCR-Verarbeitung) sind die Inhalte damit unter Umständen für eine computergestützte Analyse wertlos – die Dokumente gehen für die Beweisführung verloren.

Der im Projekt K.Rex vorgestellte Ansatz soll die menschliche Wahrnehmung der Dokumenten-Anmutung in selbstlernenden Verfahren simulieren und für die Interpretation von Dokumenten nutzbar machen. Text- und bilderschließende Verfahren werden multimodal kombiniert, um die Treffergenauigkeit bei der semantisch reichhaltigen Ableitung von Sachverhalten signifikant zu erhöhen und fehlende oder fehlerhafte textuelle Inhalte zu kompensieren. Effiziente Annotation formalisiert das Wissen der Ermittlerinnen und Ermittler, wobei fallspezifischen Besonderheiten und sich verändernden Betrugsmustern Rechnung getragen wird. Es werden adaptive Lernverfahren erforscht, die eine effiziente Anwendung auf neue Fall-Modelle bei minimalem zusätzlichem Annotationsaufwand ermöglichen. Derartige computergestützte Methoden beeinflussen den Ermittlungsprozess im soziotechnischen Gefüge maßgeblich. Daher werden auch die Aspekte Vertrauen, Verantwortung und Transparenz im Projekt ausführlich behandelt, sowie die Handlungsspielräume der Akteure und die notwendigen ethischen Rahmenbedingungen ausgelotet. Eine begleitende Labor-Evaluierung sorgt für bedarfsgerechte Ausrichtung der Forschung und ihrer Überprüfbarkeit.

Das Vorhaben hat das Potential, den Wirkungsgrad zukünftiger Analyse-Werkzeuge signifikant zu heben. Als Resultat wird eine massive Verkürzung der Falldurchlaufzeiten bei qualitativer Steigerung der Ermittlungsergebnisse erwartet und von den Bedarfsträgern gleichzeitig ein großer volkswirtschaftlicher Nutzen prognostiziert. In Kombination mit der Übertragbarkeit auf andere Anwendungsfelder ergibt sich – nach der Entwicklung zur Produktreife – ein entsprechend hohes Marktpotential

im In- und Ausland.

K.Rex adressiert vor allem den inhaltlichen Schwerpunkt 8 der Ausschreibung. Doch auch andere Schwerpunkte können von den erforschten Methoden zur Schaffung einer semantisch reichhaltigen, harmonisierten Datengrundlage profitieren. Im Konsortium sind die wesentlichen österreichischen Organisationen der Strafverfolgung und Kriminalitätsbekämpfung, hochkarätige wissenschaftliche Partner aus Technik und Sozialwissenschaften sowie ein Unternehmenspartner mit langjähriger Erfahrung im Anwendungsbereich vertreten.

## **Abstract**

Investigators in (criminal) prosecution have to sift through, inspect and evaluate huge amounts of heterogeneous data and documents – evidence seized while searching the homes or office premises of suspects. Due to the ever increasing amount of data used, stored and communicated daily, the need for computational means of analytical support is pressing. However, computational tools for textual analysis completely lack the intuitive ability of drawing conclusions from the structure of documents, optical characteristics, image elements (like logos, stamps, or written additions) and their positions, that comes naturally to humans. This inadequacy becomes even more pronounced considering the often defective text produced by OCR-software on documents of bad scan-quality.

K.Rex proposes an approach to capture elements of human perception in machine learned models to facilitate the interpretation of documents that otherwise cannot be considered in computer aided investigations and hence are lost from the chain of evidence. Methods for analysing text and images are combined in a multimodal system in order to increase the accuracy and pertinence of the computationally derived (possible) facts. In order to train the components of the system expert knowledge is efficiently recorded and formalised, specifically accounting for the idiosyncratic intricacies of each case and the ever changing patterns of fraud. Techniques for adaptive, dynamic learning will be explored to ensure fast and straightforward adaptation to new requirements with minimal effort for additional manual annotation.

The socio-technical implications of computer-aided investigations are substantial. Aspects of trust, accountability and transparency will be discussed and an ethical framework established. Early integration of sociological expertise can counteract and prevent effects like cognitive bias, preconceptions, or prejudice unintentionally instigated by the software. Iterative evaluation in a laboratory environment throughout the entire process ensures that the research meets the demands and expectations of the target audience.

The project has the potential to improve the performance and efficiency of future analysis tools significantly, resulting in a massive reduction in turnaround times for the prosecution while at the same time improving the quality of the investigative results. After developing the product to the required level of maturity, the predicted economic benefits are immense. This yields a huge market potential, both nationally and internationally – even more so due to the high relevance for various other domains.

K.Rex mainly addresses the focal points of topic 8 of the call. However, methods for creating, harmonizing, and integrating semantically rich datasets are highly pertinent to other focus topics of the call.

The consortium comprises the essential institutions concerned with prosecution and crime control in Austria, as well as top-level technical and social science research institutes and an SME with years of experience pertinent to the field.

## **Projektkoordinator**

**m2n - consulting and development gmbh**

## **Projektpartner**

**Finanzmarktaufsichtsbehörde**

**Bundesministerium für Finanzen**

**Bundesministerium für Verfassung, Reformen, Deregulierung und Justiz**

**Research Studios Austria Forschungsgesellschaft mbH**

**Universität Wien**

**Bundesministerium für Inneres (BMI)**

**Technische Universität Wien**