



NEWS EE



A Digital Investigator for
Historical Newspapers



In diesem Projekt arbeiten gemeinsam
Geisteswissenschaftler*innen,
Informatiker*innen und Bibliothekar*innen
an historischen Zeitungen
aus drei Nationalbibliotheken
(Österreich, Finnland, Frankreich)

 <https://www.univie.ac.at/newseye/>

 www.newseye.eu

 @newseyeeu

Automatische Text- erkennung

Kontakt:

Roger Labahn

✉ roger.labahn@uni-rostock.de



Vom Tage.

Im niederösterreichischen Landtage wurde heute wieder über die Erhöhung der Spitalsgebühren nicht referirt, obwohl die festgesetzte Frist um vier volle Tage überschritten ist. Die Subcomités bewähren sich.

Die Automatische Texterkennung liefert eine Transkription für jede Textzeile

Die automatische Texterkennung (Automated Text Recognition, ATR) hat die Aufgabe, die in Bildern enthaltenen Textinhalte automatisch zu transkribieren. In diesem grundlegenden Schritt des Workflows werden die Textdaten extrahiert, welche in darauf folgenden Arbeitsschritten bei der Verarbeitung natürlicher Sprachelemente Anwendung finden.

In unserem Fall wenden wir Modelle auf segmentierte Textzeilenbilder an, die das Ergebnis eines Layoutanalyse-Algorithmus sind. Diese Modelle umfassen tiefe künstliche neuronale Netze, die für Sequenz-Labeling-Aufgaben entwickelt wurden. Zusätzlich enthalten die Ausgaben solcher ATR-Modelle Schätzungen für Wahrscheinlichkeiten der einzelnen Zeichen, die zur Suche in der Transkription mittels Keyword-Spotting verwendet werden können.