



A Digital Investigator for Historical Newspapers

In this project, humanities researchers, computer scientists, and librarians work together on historical newspapers of three national libraries (Austria, Finland, France)

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

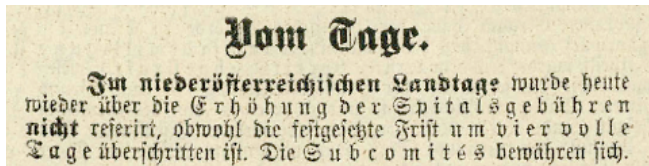
 www.newseye.eu  [@newseyeeu](https://twitter.com/newseyeeu)

Automated Text Recognition

Contact:

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.

Automatic Text Recognition provides a transcription for each text line

Automatic Text Recognition (ATR) is the task of automatically transcribing the textual contents contained in images. This is a fundamental step in the NewsEye workflow, as it provides the textual data which is needed to further be processed in Natural Language Processing tasks.

In our case, we apply models on segmented text line images, which are the results of a Layout Analysis algorithm. These models involve deep Artificial Neural Networks designed for sequence labeling tasks. Additionally, the results of such ATR models include posterior character probabilities, which can be used to search in the transcription by means of keyword spotting.