Klyueva Olga
CIS, LMU München
12.06.1917

Computerlinguistisches Arbeiten: Protokoll zum Repetitorium der Computerlinguistik.

The representation was about application, called Gini, developed by the startup company, based in Munich. The application is an Information Extraction System, which helps to extract important for the user information from letters, invoices, bills etc. The work of the Application was presented on the example of the insurance letter for paying the bill. As a result, IBAN, amount of money and the name of the person were extracted from the letter. The application can receive as an input PDF, scanned PDF, JPG document formats. After Computer Vision software comes OCR to identify the text. The information extraction is made with the help of CRF and regular expression technique.

Until now it is managed to extract over 50 sorts of information, among them are bank information, payment details, addresses, amounts, tax. IDs etc.

At the end of the meeting the future plans were mentioned. It is planned to work with deep machine learning. Potential areas for this are in getting more and smart data (dealing with feedbacks), Computer Vision, OCR engine.