

## PoCoTo – Ein Werkzeug zur interaktiven Nachkorrektur

Das am Centrum für Informations- und Sprachverarbeitung der LMU München entwickelte Nachkorrekturwerkzeug PoCoTo (Post Correction Tool) unterstützt eine interaktive Nachkorrektur von OCR-Ergebnissen. Das Werkzeug ermöglicht es seinen Benutzern sowohl einzelne Wörter, als auch ganze Fehlerserien zu korrigieren.

Über ein an PoCoTo angebundenes Sprachprofilierungswerkzeug (Profiler) können vermutete OCR-Fehler im Dokument identifiziert werden. Der Profiler verwendet dazu eine Liste von historischen Schreibmustern mit einem modernen Lexikon, um zwischen echten OCR-Fehlern und historischen Schreibvarianten im Dokument zu unterscheiden. Dabei produziert der Profiler Korrekturvorschläge unter Berücksichtigung historischer Varianten moderner Wörter, die wiederum bei der Nachkorrektur in PoCoTo zur semiautomatischen Korrektur von Wörtern und Fehlerserien verwendet werden können.

Florian Fink

Centrum für Informations- und Sprachverarbeitung (CIS)

Ludwig-Maximilians-Universität München