

PAGE-XML

Klug, Helmut W.; helmut.klug@uni-graz.at

Die PAGE-XML-Formate werden verwendet, um den Seiteninhalt von Quelldokumenten, die als Bilddigitalisate vorliegen, zu beschreiben und über Bildkoordinaten in den Digitalisaten zu verorten. Die Daten umfassen Layout, Textinhalt und mögliche Informationen zu Bildbearbeitung (*dewarping*, *deskewing*) sowie Ground-Truth-Daten, die eine Beurteilung der Ergebnisse von automatisierten Erkennungsroutinen zulassen, bei denen PAGE-XML gerne als Output verwendet wird.

Literatur:

- Pletschacher, S.; Antonacopoulos, A.: The PAGE (Page Analysis and Ground-truth Elements) Format Framework. In: Proceedings of the 20th International Conference on Pattern Recognition (ICPR2010), Istanbul, Turkey, August 23-26, 2010: 2010, S. 257-260.

Software:

T-Pen, Transkribus

Verweise:

Topografische Edition, XML, Digitalisierung, HTR, OCR

Projekte:

PAGE-XML auf Git-Hub, Pattern Recognition amp; Image Analysis Research Lab der Universität Sanfort, Manchester

Themen:

Digitalisierung

Zitiervorschlag:

Klug, Helmut W.. 2021. PAGE-XML. In: KONDE Weißbuch. Hrsg. v. Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt "Kompetenznetzwerk Digitale Edition". URL: <https://gams.uni-graz.at/o:konde.154>