

Ergebnisse: Ground-Truth-Daten

Erstellt von: Marcus Baumgarten (Herzog August Bibliothek)

Für die Workflow-Tests und die Erstellung von Ground-Truth-Daten wurden mehrere frühneuzeitliche Texte verwendet. Dabei wurden nur deutschsprachige Werke in Fraktur-Schriftart herangezogen. Insgesamt wurden ca. 1500 Textseiten volltextdigitalisiert, wovon wiederum - aus Zeitgründen - nur eine Teilmenge zur Eruierung benutzt wurden. Die erzeugten page.xml-Dateien wurden mit Hilfe der Imagedigitalisate manuell korrigiert und anschließend ein automatisierter Abgleich mit Dinglehopper erstellt und als html-Seite gespeichert. Im Einzelnen waren das:

Bildlink	Beschreibung	CER	WER
Triumpf 00813	Fraktur; große Initialen, unterschiedl. Schriftgröße, leichter Widerdruck	0.0063	0.0122
Triumpf 00827	Fraktur; Prosa und Lyrik gemischt	0.0642	0.0678
Triumpf 00831	Einfacher Blocksatz in Fraktur	0.0085	0.0334
Triumpf 00836	Gemischte Schrifttypen (Fraktur und Antiqua), Hervorhebungen (Kursivierung, Sperrung, Faltung), häufige Verwendung von Majuskeln	0.0093	0.0662
Abschied 00289	Fraktur; einfacher Blocksatz; Kursive in Antiqua	0.0076	0.038
Abschied 00297	Fraktur; einfacher Blocksatz; Kursive in Antiqua	0.0219	0.0588
Abschied 00304	Fraktur; unterschiedlicher Satzspiegel; viele Eigennamen (Personen)	0.0601	0.0778
Abschied 00316		0.0477	0.0993
Relation 00121	Titelei; Fraktur	0.045	0.0323
Relation 00126	Fraktur; Versch. Schriftgröße, Ziffern	0.1055	0.0893
Relation 00146	Fraktur, vermischte	0.04	0.065

	Schrifttypen; Marginalien		
Relation 00184	Fraktur; vermischte Schrifttypen	0.0664	0.0455