

Dokumentenmanagement

Dr. Wolfgang Schade
Gesellschaft zur Förderung angewandter Informatik e.V. (GFal)
Volmerstr. 3, 12489 Berlin
Tel.: 030 814 563 470, Fax: 030 814 563 302
E-Mail: schade@gfai.de, Internet: www.gfai.de

Auf dem Stand werden Software-Lösungen der GFal zur Erfassung von Dokumenten und zur nachträglichen Bearbeitung von Scan-Images vorgestellt:

1. Ein Programmsystem zur Unterstützung der interaktiven intelligenten Erfassung von Zeitungsartikeln

Damit kann ausgeführt werden:

Anlage des Ordners:

Themengebiet und Signatur des entsprechenden Ordners

Artikelerfassung

Namenserfassung (Vorname, Nachname, Vorsatz (Graf..), Nachsatz (von..), Titel (Dr.))

Sprachklassifizierung des Artikels (deutsch, polnisch, tschechisch...)

Anzeige des Scan-Images mit Zoomfunktion

Textausrichtung (bei schräg aufgeklebten Artikeln)

Interaktive Textbereichsseparierung

Interaktive Abbildungseparierung

Abbildungsbeschreibung (Karikatur, Foto, Skizze)

Zuordnung von Bildunterschriften

Bildinhaltsbeschreibung, falls notwendig

Erfassung von Autoren und Fotografen

Erfassung der Artikelüberschrift(en)

Artikelklassifizierung (Anzeige, Gedicht, Reportage, Interview, Roman)

Kennzeichnung, ob Artikel/und oder Abbildung(en) freigegeben werden können

gesonderte Erfassung des Zeitungsnamens (Zuhilfenahme eines Scroll-Feldes)

gesonderte Erfassung des Erscheinungsdatums

mit entsprechenden Eintragungen in eine MySQL-Datenbank.

Das Interface bietet außerdem die Möglichkeit, Ergänzungen hinzuzufügen, wie z.B. persönliche Daten der Autoren und PND, und ermöglicht die Kontrolle/Korrektur der durch die eingebundene OCR gelieferten Ergebnisse.

(s. dazu auch den Vortrag "Erfahrungen bei der Erfassung eines Zeitungsarchivs")

2. Findex- Ein Programmsystem zur Indexierung von digitalisierten Büchern

Ausgehend vom Buchindex oder von einer eingegebenen Schlagwortliste werden die OCR-Ergebnisse der Seiten durchsucht, wobei zum Auffinden die Methode der unscharfen Suche eingesetzt wird.

3. Programme zur nachträglichen Bearbeitung eingescannter Farbbimages,

zum Beispiel zur Farboptimierung bei nicht optimal eingestellten Scannern.