DaCaPo: Ein System zur Inhaltserfassung von Zeitungen

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

Das Erfassungssystem DaCaPo ist eine Client-Server-Anwendung zur teilautomatisierten Inhaltserfassung von Dokumenten.

Mit dem Programmsystem zur interaktiven intelligenten Inhaltserfassung von Zeitungsartikeln aus deren Seiten-Images kann ausgeführt werden:

1. Anlage des Ordners:
   Themengebiet und Signatur des entsprechenden Ordners

2. 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 Abbildungsseparierung
   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

3. Eintragung der Ergebnisse in eine MySQL-Datenbank

4. Das Interface bietet außerdem die Möglichkeit, Ergänzungen hinzuzufügen, wie z. B persönliche Daten der Autoren und PND, und die durch die eingebundene OCR gelieferten Ergebnisse zu kontrollieren bzw. zu korrigieren.

5. **Neu ist unser auf dem Server zusätzlich laufendes Entwicklungstool, mit dem eine automatische Bereichsseparierung, Stempelidentifizierung und eine partielle Handschrifterkennung (Numerik) möglich sind. Diese automatisch erhaltenen Resultate können nach DaCaPo übernommen werden.**

Die in der MySQL-Datenbank abgelegten Resultate lassen sich sowohl hausintern wie auch für Internet-Präsentationen nutzen.