

Rationelles Erstellen eines Archives aus Zeitungen und Zeitschriften, Büchern und anderen Dokumenten

PaperIn DocClip mit OCR (Optical Character Recognition) ermöglicht die Textsuche im Volltext- sowie SQL-fähigen Datenbanken

The screenshot shows the PaperIn DocClip software interface. The main window displays a scanned document page with the following content:

Der schweizerische „Tierfreund“ und das Geflügelschächten.

(L.) Der „Tierfreund“, das Organ der deutsch-schweizerischen Tierschutzvereine, schließt das Jahr 1908 mit einem Artikel über das bundesgerichtliche Urteil betreffend das Geflügelschächten. Er bringt einen Auszug aus einer diesbezüglichen Separatarbeit der „Leipziger Flugschriftenammlung“ zur Betäubungsfrage der Schlachttiere und verweist die Interessenten auf den ganzen Artikel, der beim Vorstandspräsidenten der Tierschutzvereine zu haben ist. Uns genügt, wie dem „Tierfreund“, der Auszug.

Das Bundesgericht hatte in **kurzem** so erwogen: Die Bundesverfassung garantiert die Kultusfreiheit. Das Schächten, d. h. das Schlachten der Tiere ohne vorhergehende

The left sidebar shows the following segments:

- Artikelinfo:
 - Title: Israelisches Wochenblatt..
 - Author(s): Autor...
 - Theme: Schweizerische Tierfreund
- Erstellen, Bearbeiten, Löschen buttons
- Artikel:
 - israelisches Wochenblatt..: Autor...: Schächten: 01.01.
- Erstellen, Bearbeiten, Löschen buttons
- Artikelsegmente:
 - 2: 9: Autosegment
 - 2: 11: Autosegment
- Seitensegmente:
 - 2: 1: Autosegment
 - 2: 2: Autosegment
 - 2: 3: Autosegment
 - 2: 4: Autosegment
 - 2: 5: Autosegment
 - 2: 6: Autosegment
 - 2: 7: Autosegment
 - 2: 8: Autosegment
 - 2: 9: Autosegment
 - 2: 10: Autosegment
 - 2: 11: Autosegment

The bottom of the window shows a search result snippet:

(*) Der „Tierfreund“, das organ der deutsch-schweizerischen Tierschutzvereine, schliesst das Jahr 1908 mit einem Artikel über das bundesgerichtliche Urteil betreffend das Geflügelschächten. Er bringt einen Auszug aus einer diesbezüglichen Separatarbeit der „Leipziger Flugschriftenammlung“ zur Betäubungsfrage der Schlachttiere und verweist die Interessenten auf den ganzen Artikel, der beim Vorstandspräsidenten der Tierschutzvereine zu haben ist. Uns genügt, wie dem „Tierfreund“, der Auszug. Das Bundesgericht hatte in kurzem so erwogen: Die Bundesverfassung garantiert die Kultusfreiheit. Das Schächten, d. h. das Schlachten der Tiere ohne vorhergehende

PaperIn DocClip...

- scannt Dokumente, assembliert Artikel welche textbezogen zum Image referenziert werden.
- segmentiert und klassifiziert die Seiten in Text- und Bildabschnitte.
- ermöglicht das zusätzliche Einrichten von Indexfeldern.
- ermöglicht die Erkennung des Textes (OCR) aus gewählten (Bild)Segmenten.
- ermöglicht das Auslesen von Texten der meisten westeuropäischen Sprachen sowie von Sprachen in kyrillischer Schrift.

- ermöglicht das Validieren des ausgelesenen Textes und dessen Korrektur mit Hilfe von speziellen Wörterbüchern.
- erzeugt Textdaten in verschiedenen Zielformaten, von denen einige die visuelle Referenzierung vom erkannten Text auf das gescannte Seitenbild ermöglichen.
- führt zu Kontrollzwecken ein Protokoll, in welchem Benutzer, Datum, Zeit aufgelistet sind.
- erlaubt den Export bzw. die Bereitstellung von Textdateien für die Übernahme in Datenbanken.

PaperIn DocClip...

Basis Modul: Scannen, Segmentieren, OCR, strukturierte Textausgabe (ASCII, XML ...)

Optionale Module:

- PDF Ausgabe
- Textvalidierung mit Hilfe von Sprachlexika (CIS)
- Linguistische Analyse zur Unterscheidung von Doppeldeutigkeiten *
- Redlining (Anker und Hyperlink Editor)*
- Gothische Schrifterkennung *
- Arabische Oberflächenführung und Segmentierung *

* in Arbeit

Technische Spezifikationen:

Schrifterkennung/Schriftarten: Druckschrift (Omnifont)

Schriftgröße: 8-72 Punkt bei minimum 200 dpi

Lesebereich: Segmentierte Felder: graphische menugesteuerte Oberfläche

Bildspeicherung: in verschiedenen Auflösungen und Formaten, z.B. CCITT G3/G4

Datenausgabeformat Text: ASCII, ISO, Xdoc²⁾, PDF³⁾, XML

Für Datenbank: Im Text ausserhalb liegendes Bild referenziert oder in Datei eingebettet.

Image Akzeptanz: verschiedene Auflösungen und bilevel Formate, z.B. CCITT G3/G4, GIF, JPEG

Betriebssystem: NT 4.SP6 ¹⁾, Windows 2000¹⁾, Windows XP¹⁾

PC Voraussetzungen: Intel Pentium III, min. 1GHz, min. 256 MB RAM, minimum SuperVGA; 19" Bildschirm empfohlen

Scanner:

Fujitsu, und viele andere TWAIN kompatible Scanner

Hersteller:

ARPA Data GmbH
Albisstrasse 36

CH 8134 Adliswil

Tel: +41 1 709 09 79

¹⁾Registriertes Warenzeichen von Microsoft™

²⁾Registriertes Warenzeichen von Scansoft Inc

³⁾Registriertes Warenzeichen von Adobe Systems Inc.