

PAPERIN

Multifunctional Data Entry System

DocClip

Creating an archive with news papers and magazines, books and other documents

PaperIn DocClip with OCR (Optical Character Recognition) allows searching in full text retrieval databases as well as sql based databases

The screenshot displays the PaperIn DocClip software interface. The main window shows a news article titled "Weihnachtsfilme für jeden Geschmack". The article text is partially visible, discussing the Christmas movie program in Basel. The interface includes a menu bar (File, View, Edit, Options, Special, Admin, Language, Help), a toolbar with various icons, and a sidebar on the left with a file tree structure. The article text is highlighted with red boxes, and several numbered callouts (1-8) are placed over the text and images. The callouts point to specific parts of the article, such as the title, the first paragraph, and various paragraphs of text. The article text includes: "1. Weihnachtsfilme für jeden Geschmack", "2. Effektives Weihnachtsprogramm bieten die Basler Kinos zwar nicht mehr, aber den obligaten Walt-Disney-Zeichentrickfilm gibt es nach wie vor. Und auch andere Filme scheinen speziell auf die Festtage hin ins Kino zu kommen.", "3. Die Basler Kinos bieten die Festtage kein Weihnachtsprogramm im herkömmlichen Sinn. Dies ergab eine Umfrage bei den drei wichtigsten Basler Kinobetreibern. Dennoch laufen derzeit einige Filme, die als typische Weihnachtsfilme gelten. Allen voran der Walt-Disney-Zeichentrickfilm: 1994 war es der «Lion King», der die Kinder in Scharen anzog. Letztes Jahr erfreute «Pocahontas» die Kleinen und dieses Jahr ist es «The Hunchback of Notre Dame» (Hollywood). Heuer schliesst sich auch das Kinderkino Atelier dieser Zeichentrickfilmtradition an. Es zeigt von 24. Dezember bis zum 2. Januar täglich", "4. mal den Film «Der blaue ...», sozusagen die Weihnachtsgeschichte auf italienisch. Denn hier bringt Befana, zugleich Hexe und Gräfin, den Kindern am Dreikönigstag Geschenke. Der Film von Enzo J. Alò lief zwar schon im November für kurze Zeit, wurde dann aber zurückgezogen, um ihn gezielt über die Festtage nochmals zu zeigen. Einen weiteren Kinderfilm neben den erwähnten Zeichentrickfilmen bietet das Rex mit «Napoleon». Hier meistert ein Hundewelppe die Gefahren der australischen Wildnis. Die Kinobetreiber haben aber nicht nur an die Kinder", "5. n auch an die Erwachsene gedacht. Der soeben angeaufene Schwarzenegger-Film «Jingle All die Way» (Küchlin) kommt auf jeden Fall zur rechten Zeit in die Kinos, hat er doch Weihnachten zum Thema und dazu einen zugkräftigen Namen, der die ganze Familie anzieht. Liebhaber von Actionfilmen werden ebenfalls auf ihre Rechnung kommen: Sylvester Stallone rettet in Rob Cohens Streifen «Daylight» die nach einer Explosion Verschlütteten aus einem Tunnel (Capitol). Gute Unterhaltung verspricht «The First Wives Club» (Eldorado). Der Club der Ex-Frauen, gespielt von Bette Midler, Diane", "6. i und Goldie Hawn, zeigt Ex-Gatten, wo's langgeht. Die Betreiberin der Kinos Atelier, Camera, Club und Movie, die Studiokino AG, hat für anspruchsvollere Kinogänger vorgesorgt und setzt auf die bewährte Mischung von Studio- und Dokumentarfilmen. Dass es kein Weihnachtsprogramm im herkömmlichen Sinne mehr gibt, hat einerseits damit zu tun, dass das Publikum kein Interesse mehr hat an sogenannten «Heile Welt»-Filmen, andererseits aber auch mit den Filmverleihern. Edgar Müller, Geschäftsführer der Kinos Capitol, Central, Rex und Scala, betont, dass die Programmation stark von den Filmverleihern abhängig ist, die den Kinobetreibern wenig Spielraum lassen. Gegen eventuell aufkommende Langeweile oder Weihnachtskoller ist folglich vorgesorgt, denn die meisten Kinos sind auch über die Festtage geöffnet. Einzig am Heiligabend ist ein Grossteil der Spielstätten geschlossen. Dann ist der Besucherzuspruch in der Regel so klein, dass sich der Betrieb einfach nicht lohnt. Generell nimmt die Besucherzahl über die Festtage aber deutlich zu. Für Manfred Thurov von der Firma Walch Kinobetriebe AG (Eldorado, Hollywood, Küchlin und Plaza) hat dies aber weniger mit der Tatsache zu tun, dass die Kinder Schulferien haben. Die Kinokassen dürften auf jeden Fall – egal ob Weihnachts- oder Ferienbonus – Ende Monat stimmen.

PaperIn DocClip...

- scans documents, assembles articles which are referenced tet related to the image.
- segments and classifies pages in text and image parts.
- allows the additional definition of index fields.
- allows the recognition of text (OCR) from chosen image segments.
- allows the recognition of most West European languages as well as Cyrillic.
- allows the validation of the recognized text and its correction with help of special dictionaries.
- creates text files in different output formats where some of them allows a visual referencing of the recognized text on the scanned image.
- writes a protocol for controlling purposes with users, date, time.
- allows the export of text files for loading directly into databases.

PaperIn DocClip...

Basic module: Scanning, segmenting, OCR, structured text output (ASCII, XML ...)

Optionale modules:

- PDF output
- Validation of text with help of language lexica (CIS)
- Linguistic analysis to distinguish between different meanings *
- Redlining (ancor and hyperlink editor)*
- Gothic font recognition *
- Arabic screen layout and segmentation *

* in work

Technical specifications:

Character recognition/fonts: printed fonts (Omnifont)

Character size: 8-72 dots at minimum 200 dpi

Area: Segmented fields: graphical menu driven screen layout

Image saving: Various resolutions and formats, e.g. CCITT G3/G4

Data output for text: ASCII, ISO, Xdoc²⁾, PDF³⁾, XML

For databases: Reference of image to outside area of database or embedded in file.

Image acceptance: different resolutions and bilevel formats, e.g. CCITT G3/G4, GIF, JPEG

Operating system: NT 4.SP6 ¹⁾, Windows 2000¹⁾, Windows XP¹⁾

PC requirements: Intel Pentium III, min. 1GHz, min. 256 MB RAM, minimum SuperVGA; 19" monitor recommended

Scanner:

Fujitsu, and other TWAIN compatible scanner

Producer:

ARPA Data GmbH
Albisstrasse 36

CH 8134 Adliswil

Tel: +41 1 709 09 79

¹⁾Registered trademark of Microsoft™

²⁾Registered trademark of Scansoft Inc

³⁾Registered trademark of Adobe Systems Inc.