

## **OCR-Texterkennungsprogramme im Test**

### **Transym, OmniPage und FineReader überzeugen im Zusammenspiel mit der Archivierungslösung "Office Manager"**

Königs Wusterhausen - Sollen gescannte Dokumente in einem digitalen Archivsystem per Volltextsuche wiederauffindbar sein, dann ist neben der Archivierungssoftware auch ein gutes Texterkennungsprogramm erforderlich. Denn Papier-Dokumente liegen nach dem Scan nur als grafisches Abbild vor - der enthaltene Text muss für die Datenverarbeitung noch ermittelt werden. Moderne Softwarelösungen verwenden intelligente Optical Character Recognition-Technologien, um die einzelnen Zeichen im Bildmuster zu erkennen. Diese werden zu Wörtern zusammengefasst und mit Wörterbüchern abgeglichen, Falscherkennungen können dann automatisch korrigiert werden.

Doch bei der Erkennungsquote trennt sich die Spreu vom Weizen, wie ein aktueller OCR-Test der DMS-Spezialisten vom Softwarebüro Krekeler zeigt. Sie stellten die Texterkennungsprogramme

- Nuance OmniPage (Versionen 17.1 und 18.0),
- ABBYY FineReader for ScanSnap 4.1,
- Microsoft Office Document Imaging 2003,
- Transym 3.0.2.4 und die
- Open-Source-Lösung Tesseract 3.0

im Zusammenspiel mit ihrer etablierten DMS-Lösung Office Manager auf den Prüfstand. Letztere verfügt über entsprechende Schnittstellen für die Ankopplung der OCR-Programme.

"Zunächst haben wir die Lösungen an einem Testdokument mit sauberem, schwarzem Druck auf weißem Hintergrund angewandt. Die zweite Vorlage war da schon anspruchsvoller: Die auf der Rückseite eines Kontoauszuges gedruckten AGB eines Kreditinstitutes mit vorhandenen Flecken und besonders kleiner Schrift. Die aufgetretenen Fehler wurden festgehalten und die Erkennungsquote ermittelt", erläutert Harald Krekeler, Inhaber des gleichnamigen Softwarebüros, den Testablauf. Während sich alle Kandidaten beim ersten Testdokument nahezu gleich gut schlugen, zeigten sich bei den AGB zum Teil deutliche Unterschiede hinsichtlich der Erkennungsgenauigkeit.

Mit einer Erkennungsquote von 99,0 Prozent schnitt Transym beim Vergleichstest am besten ab. OmniPage sowie der ABBYY FineReader belegten mit 98,7 Prozent beziehungsweise 98,3 Prozent den zweiten und dritten Platz. Mit merklichem Abstand folgen Microsoft Office Document Imaging (95,3 %) sowie die einzige Open-Source-Lösung im Test - Tesseract (92,9 %). "Das hervorragende Ergebnis von Transym hat uns überrascht, während die Erkennungsquoten von OmniPage und FineReader den Erwartungen entsprachen. Das schlechte Abschneiden von Tesseract hingegen liegt darin, dass die Lösung nur auf die reine Zeichenerkennung, jedoch dabei von Hause aus nicht auf ein Wörterbuch zur Plausibilitätsprüfung zurückgreift", resümiert Harald Krekeler.

[Zum OCR-Vergleich](#)

## **Über das Softwarebüro Krekeler**

Das Softwarebüro Krekeler beschäftigt sich seit 1996 intensiv mit dem Thema Dokumentenmanagement und -archivierung und hat sich auf die Entwicklung entsprechender Business-Lösungen spezialisiert. Auf der ITK-Fachmesse CeBIT wurde 1997 erstmals die DMS-Lösung "Office Manager" vorgestellt.

Das Prinzip der Lösung war und ist denkbar einfach: Papierdokumente werden gescannt, mittels der Volltextindizierung erfasst und mit beliebig vielen Attributen katalogisiert. Gemeinsam mit elektronischen Dokumenten nahezu jeden Formates gelangen sie in ein zentrales Datenverzeichnis. Aus diesem lassen sich die Dokumente mit Hilfe der zugeordneten Attribute und der Volltextsuche mitarbeiterübergreifend abrufen, übersichtlich auflisten und verarbeiten.

Das über Jahre fortgeführte und optimierte Ergebnis, der heutige "Office Manager", ist inzwischen auch als leistungsfähige Version "Office Manager Enterprise" für große Unternehmensnetzwerke sowie neben deutscher auch in englischer und polnischer Sprache erhältlich.

### **Presse-Kontakt:**

[Softwarebüro Krekeler](#)

Lösungen für Dokumentenmanagement

Friedrich-Engels-Str. 45

D-15712 Königs Wusterhausen

**PresseDesk**

Tel: +49 30 20095130

E-Mail: [softwarebuero-krekeler@pressedesk.de](mailto:softwarebuero-krekeler@pressedesk.de)