

Daten digitalisieren

Thomas Engel (Foto: zvg)

Immer mehr Praxen stellen auf digitale Systeme um, die Röntgen werden nur noch digital erstellt, die Fotos sind schon lange nur noch als jpg oder tiff vorhanden, und immer mehr Gipsmodelle werden durch Bits und Bytes ersetzt. Nur die Krankengeschichte hat den Weg in die digitale Zukunft noch nicht vollends gefunden. Die seit Jahrzehnten lieb gewordenen KG verbleiben in alter Form und schrecken viele Zahnärzte davon ab, voll in die neue digitale Zukunft zu starten.

Ist der grosse Schritt in die neue Welt dann doch vollbracht, stellt sich rasch eine entscheidende Frage: Soll mit zwei Systemen (analog und digital) parallel oder doch nur noch elektronisch weitergearbeitet werden? Heute sind bereits viele Programme auf dem Markt, welche das Verwalten der unterschiedlichsten Patientendaten recht einfach und übersichtlich ermöglichen. Klickt der Zahnarzt einen Patienten an, erscheinen auf dem Bildschirm alle wichtigen Daten dieses Patienten in geordneter Form.

Alles wunderbar, wäre da nicht noch die alte KG, welche vergilbt und irgendwie unpassend neben dem Behandlungsstuhl liegt und kaum mehr Beachtung findet. Genau hier möchte ich mit meinem neuen Thema ansetzen: die alten KG digitalisieren! Bereits recht günstig gibt es heute gute Scanner auf dem Markt, so etwa der Canon DR-4010C, Epson GT-S50N, Canon DR-5010C, Fujitsu ScanSnap S1500, Canon ScanFront 220/P, Avision AV122, Fujitsu fi-6140 oder der Fujitsu ScanSnap, um nur einige zu nennen. Oft können diese Geräte doppelseitig scannen und je nach Modell und Auflösung bis zu 100 Seiten pro Minute abarbeiten. Die Daten können in verschiedenen Formaten gespeichert werden, so auch als PDF.

Text erkennen

Was ist OCR? Optische Zeichenerkennung (OCR) steht für den Vorgang, bei welchem aus Bildern (oder PDF, usw.) einer Seite der Text herausgezogen wird. Diese Bilder werden oft durch Scannen oder Fotografieren eines Dokuments oder via E-Mail in digitaler Form auf den Rechner gebracht.

Bilder bestehen nicht aus bearbeitbaren Textzeichen, sondern aus zahlreichen winzigen Punkten (Pixeln) oder je nach Speicherart aus Vektoren, welche zusammen ein Bild oder eben das Abbild eines Textes ergeben. Während des OCR-Vorgangs analysiert das Programm das Bild und erstellt daraus Text, der direkt am Computer bearbeitet werden kann. Er ist nun also in Form von Schriftzeichen gespeichert.

Entsprechende OCR-Programme wie etwa Abbyy Fine Reader Pro, Scan Soft Omni Page Pro, Iris Readiris Pro und viele andere mehr bieten vielseitige Erkennungsoptionen, welche je nach Vorlage sehr gute Resultate liefern. Heutige Programme erkennen nicht nur die reinen Schriftzeichen, sondern auch die Darstellung und Texteditierung, vorhandene Bilder und Tabellen, teilweise auch die Schriftart und die Schriftfarbe und nicht zuletzt auch die Sprache. Dank all diesen Informationen können diese Programme den erkannten Text nebst der reinen Texterkennung auch editieren und korrigieren und so weit aufbereiten, dass er ohne grösseren Aufwand als PDF oder Worddatei gespeichert und weiterverwendet werden kann.

Handschrift erkennen

Unter ICR (Intelligent Character Recognition) versteht man das Erkennen handschriftlich erstellter Texte. Je nach Anwendung unterscheidet man zwischen Erkennung vektorbasierter Handschrift, wie sie bei PDA (kleine elektronische Agenden) zum Einsatz kommt, oder Erkennung gerasterter Handschrift, wie sie zur Auswertung gescannter Vorlagen benötigt wird. Bei der Handschrifterkennung wird eine deutlich niedrigere Genauigkeit als bei der OCR erreicht. Viele Anwendungen zwingen den Benutzern spezielle Schriftzeichen auf, womit eine bessere Erkennung ermöglicht

wird. Da dies bei bereits erstellten Dokumenten nicht möglich ist, müssen wir vorerst mit der schlechten Erkennungsrate auskommen. Hier stellt sich momentan noch die Frage, ob der handgeschriebene Text als solcher erkannt werden muss oder ob nicht die Erstellung eines grafikbasierten PDF doch sinnvoller ist. Laut Versprechungen der Hersteller soll die Handschrifterkennung in Zukunft deutlich besser werden. Viele Programme besitzen zudem einen Trainingsmodus, bei jeder Anwendung lernt das Programm die Schriftzeichen seines Besitzers etwas besser kennen und ermöglicht so mit der Zeit eine recht gute Erkennungsrate.

Begriffe

PDF (Portable Document Format): ein plattformunabhängiges (Linux, Sun, Windows, Mac...) Dateiformat für Dokumente, das vom Unternehmen Adobe Systems entwickelt und 1993 veröffentlicht wurde. Es ermöglicht den Datenaustausch verschiedener Systeme und Programme.

OCR (Optical Character Recognition): beschreibt die automatische Texterkennung von einer analogen Vorlage.

ICR (Intelligent Character Recognition): nach Anwendung wird unter diesem Begriff die Erkennung der Handschrift oder die Plausibilitätsprüfung von OCR-Ergebnissen verstanden.

Fortsetzung folgt...

