

OCR 1994 – Quo vadis?

Andreas Herberger

Mit schöner Regelmäßigkeit wird das Thema OCR oder besser Zeichenerkennung in den PC-Zeitschriften reihum aufgegriffen. Waren es früher noch mehrjährige Intervalle in denen die Thematik bearbeitet wurde, scheint sich jetzt ein jährlich wiederkehrender Brauch zu entwickeln. Doch vielleicht gibt es ja auch in der Sache selbst begründete Notwendigkeit, sich gerade jetzt erneut mit dem Thema auseinanderzusetzen?

Und in der Tat, in den letzten 6 Monaten hat sich einiges getan im Markt der Zeichenerkennungsprogramme.

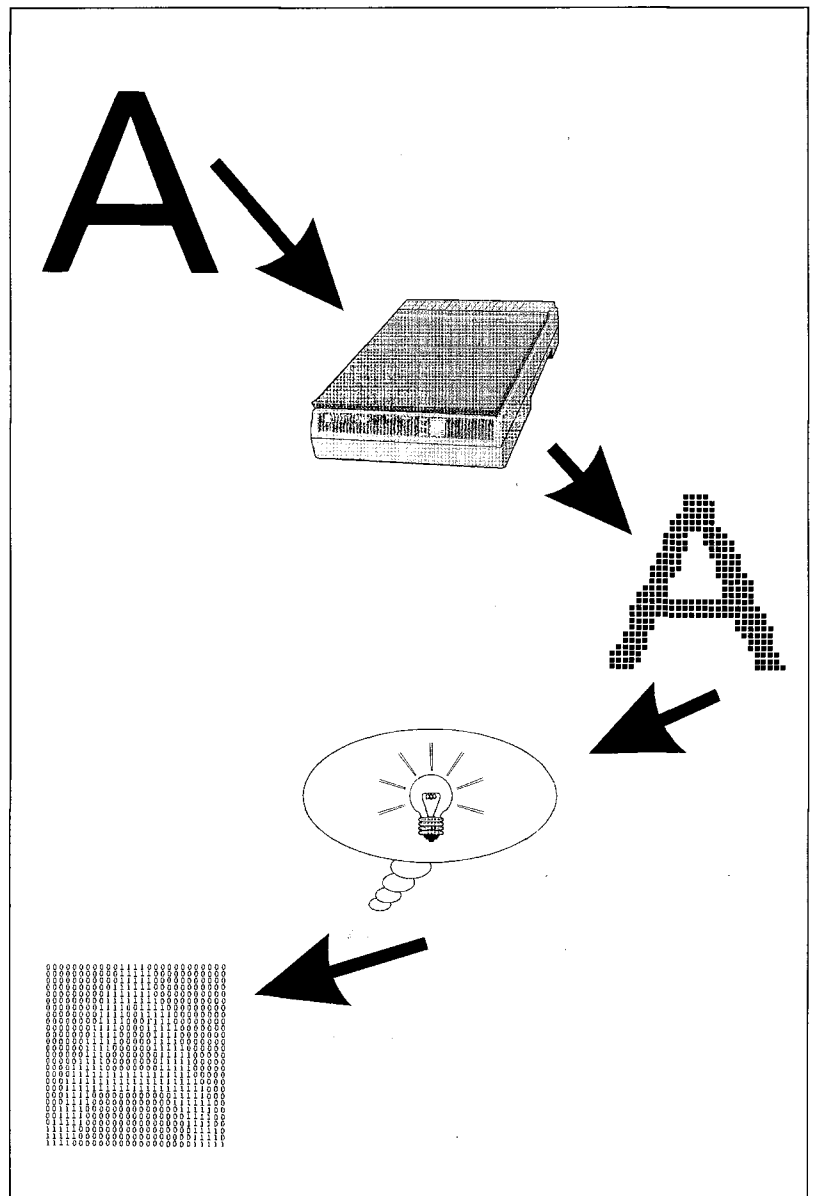
Scannerpreise und "Bundle"-OCR

Die Scannerpreise sind schon länger auf dem Weg in den Keller. Mußte man vor 2 Jahren für den 300 dpi Einstiegs-Scanner noch nahezu 2000,- DM auf die Ladentheke legen – preiswerter war allenfalls ein Handy-Scanner für rund 500,- DM zu haben – sind die Preise für 300 dpi Scanner nunmehr im Bereich um 1000,- DM angelangt, und meist bekommt man sogar schon eine Einstiegs-OCR-Software dazu. Daß die großen Handelsketten ihre im letzten Jahr begonnene massive Scannervermarktung wieder zurückgenommen haben, mag nicht zuletzt daran liegen, daß eben diese "Einstiegs"-OCR-Software in den meist allenfalls zum Erlernen des OCR-Vorganges durch den Benutzer nicht aber zum Erzielen brauchbarer OCR-Ergebnisse geeignet ist. Nur so ist es zu erklären, daß trotz aller Massenbundlings ständig neue Anbieter in den OCR-Markt drängen, die gerade auf den enttäuschten Bundling-Kunden setzen und gewissermaßen als "Zweit-OCR" gekauft werden wollen.

Neu im Markt Textbridge und OPTOPUS-Economy

Wurde der Markt in der Vergangenheit von den Anbietern der Programme Omnipage (Caere), Wordscan (Calera) und Recognita (SZKI) dominiert, hat sich das spätestens mit dem Angebot von TextBridge durch die Firma Kurzweil (=Xerox Imaging Systems) geändert. Kurzweil ist seit Beginn der 80er Jahre synonym mit Pionierentwicklungen im Bereich der

Zeichenerkennung. Mit dem Zeichenerkennungsprogramm OPTOPUS der Wiesbadener Makrolog GmbH teilen sich die High-End Kurzweil Lesesysteme (K5XXX) seit Mitte der 80er Jahre den professionellen Erfassungsmarkt. Beide Systeme waren allerdings für den Gelegenheitseinsatz (mal gerade eben ein paar Seiten scannen) mit Preisen um die 10.000,- DM zu teuer. Um so interessanter ist, daß beide High-End Anbieter die gleiche Strategie für den Massenmarkt verfolgen: Kurzweil mit dem Einstiegsprodukt "Textbridge" für weniger als 500,- DM



Andreas Herberger ist geschäftsführender Gesellschafter der Makrolog Gesellschaft für Logik- und Computeranwendungen mbH in Wiesbaden.

und Makrolog mit der OPTOPUS-Economy-Edition für um die 1000,- DM. Textbridge führt die auf automatisches Lesen ohne Training ausgerichtete Kurzweilstrategie konsequent weiter. Der Anwender legt Seiten in den Scanner, definiert in komplizierten Fällen ggf. einige Lesebereiche, und das Umsetzen läuft sodann vollautomatisch ab und liefert, je nach Vorlagenart, ein mehr oder weniger gutes Ergebnis. Eingriffsmöglichkeiten oder besondere Fähigkeiten für große Vorlagenmengen bietet das Programm kaum, unterstützt dafür aber fast sämtliche europäischen Sprachen. Demgegenüber setzt die OPTOPUS-Economy Edition nach wie vor auf Interaktion mit dem Benutzer, unterstützt diesen aber durch ein automatisches Vortraining, das den eigentlichen Optimierungsaufwand auch bei geringen Vorlagenmengen kaum ins Gewicht fallen läßt. Durch die Möglichkeit des Upgrades auf die OPTOPUS Business Edition ist zudem sichergestellt, daß ohne Programmwechsel und damit ohne zeitaufwendiges Erlernen einer neuen Software auch ein sich einstellender Produktionsbedarf abgedeckt werden kann. Durch die Interaktion mit dem Benutzer wird darüber hinaus auch in komplizierten Fällen eine sichere Erkennung ohne Verwechslungen erreicht. Beide Programme sind noch unter einem anderen Aspekt bemerkenswert: Während bisher die käuflich zu erwerbenden Programme sämtlich nur mit einem auf die parallele Schnittstelle aufzusteckenden Hardwarekopierschutz (Dongle) betrieben werden konnten, verzichten sowohl Textbridge als auch die OPTOPUS-Economy Edition auf diesen für den Anwender in jedem Falle lästigen Zusatz. Außer Kurzweil ist noch kein leistungsfähiges Programm in die Preisklasse unter 1000,- DM eingebrochen. Auch neue Anbieter wie z. B. der Cuneiform Reader von Cognitive Technology versuchen sich zunächst im höheren Preissegmenten zwischen 1.000,- DM und 2.000,- DM (von limitierten Versionen für Handy-Scan-

ner abgesehen). Ob hier die gleiche Schere zwischen Listenpreis und Straßenpreis wie bei der übrigen Standardsoftware aufgehen wird, bleibt abzuwarten.

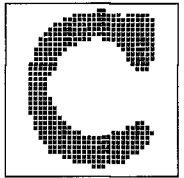
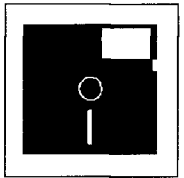
Neue Entwicklungen der Erkennungstechnik

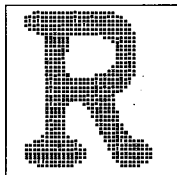
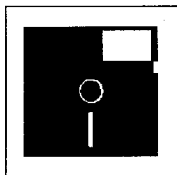
Nachdem der jahrelange Streit, ob nun die *feature-basierten Systeme* mit Wörterbuchunterstützung den trainierbaren *Pattern-matching-Systemen* überlegen sind (zu den beiden unterschiedlichen Ansätzen vgl. den Anhang "Erkennungsverfahren" auf der nächsten Seite), nach wie vor auf Entscheidung wartet, hat man sich bei den Herstellern an eine weitere Grundlagentechnologie erinnert, deren Einsatz eigentlich nahe liegt: Image Enhancement. Nachdem die HP Accupage Technologie, über deren Hintergründe von HP bisher nur sehr wenig zu erfahren war, allem Anschein nach das Image Enhancement vor der Ablieferung der Daten an die OCR-Software betreibt, hat die Firma Caere wohl als erster Softwarehersteller in ihrem neuen Omnipage Professional das Image-Enhancement scannerunabhängig eingebaut. Der Gedanke ist einfach: Normalerweise scannen OCR-Programme die Vorlagen als Schwarz/Weiß-Bilder, d. h. der Scanner liefert für jeden Bildpunkt entweder die Information Schwarz oder Weiß. Von Zwischenstufen erfährt das OCR-Programm nichts. Die Entscheidung wird alleine vom Scanner aufgrund eines fest eingestellten (Kontrast-)Schwellwertes getroffen. Oft ist dieser konstante Schwellwert aber nicht für die gesamte gescannte Seite passend. Es liegt also nahe, die Seite mit z. B. 16 Graustufen zu scannen. Das OCR-Programm kann dann in Abhängigkeit des Erkennungsergebnisses eine Kontrastanpassung durchführen. Technisch betrachtet handelt es sich also um eine Rückkoppelung des Erkennungsergebnisses auf die Kon-

trastbestimmung. Daß dies in der Vergangenheit keinen Eingang in die Programme gefunden hat, wird einsichtig, wenn man sich vorstellt, daß eine Schwarz/Weiß A4-Seite nach dem Scannen ca. 1 MB an Pixeldaten benötigt, dieselbe Seite mit 16 Graustufen aber schon auf 4 MB anwächst. Auf einem älteren Rechner mit vielleicht nur 2 MB Hauptspeicher wird sich eine solche Datenmenge nur schlecht verarbeiten lassen. Da aber heute bereits jede bessere Textverarbeitung unter Windows für produktiven Betrieb einen 80486 Rechner mit wenigstens 8 MB Hauptspeicher erwartet, dürfen auch OCR-Programme jetzt solche Anforderungen stellen. Da zu erwarten ist, daß auch die anderen Programme in Kürze diese Standardtechnik einsetzen werden, sollte beim Kauf eines Scanners nicht mehr nur auf die Auflösung, sondern auch auf die Graustufenfähigkeiten geachtet werden.

Neue Anwendungsgebiete

Die Integration von OCR-Software in Faxpakete ist mittlerweile selbstverständlich, wenn auch nach einhelliger Meinung hier noch kaum brauchbare Ergebnisse erzielt werden. Dies liegt wohl überwiegend daran, daß Faxe in Deutschland zumeist noch aus herkömmlichen Faxgeräten kommen und somit mit nur ca. 100 dpi auf der Sendeseite gescannt wurden. Zusammen mit Übertragungsproblemen sind die ankommenden "Bilder" auch für den Menschen manchmal nur schwer lesbar. Besserung ist hier in Sicht, werden doch in Zukunft eine Vielzahl von Faxen direkt aus der Textverarbeitung per Faxkarte versandt werden. Diese kommen dann in perfekter Qualität beim Empfänger an und verwenden zudem nur eine sehr beschränkte Anzahl von Schriften, so daß die Erkennung nicht schwerfällt. Warum allerdings dann nicht gleich ein Versand per Email stattfindet (der





die ASCII-Pixel Transformation beim Sender und die Pixel-ASCII Transformation beim Empfänger erübrigen würde) wird wohl noch einige Jahre ein Rätsel bleiben. Trotzdem oder gerade deswegen sind die OCR-Anbieter auf der Suche nach neuen Märkten. Eine clevere Idee, einen neuen Markt zu erschließen, ist der Einsatz von OCR-Software zum Einlesen von Visitenkarten. Mehrere Anbieter bieten hier

entweder Softwarelösungen zum Einsatz mit normalen Scannern oder sogar spezielle Visitenkartenleser an, die im Grunde nichts anderes sind als ein Bildarchivierungssystem für Visitenkarten mit integrierter OCR-Software. Die OCR-Software versucht auf der Visitenkarte die verwertbaren Informationen (wie Name und Adresse) zu erkennen und zusätzlich gleich den richtigen Feldern in einer mitgelieferten Da-

tenbank zuzuordnen. Der Schritt zum Newton mit eingebautem Visitenkartenleser, den man, anstatt die Visitenkarte entgegenzunehmen, einfach kurz darüberzieht und nach der elektronischen Aufnahme sofort wieder zurückgibt, scheint also keine Zukunftsmusik mehr. Ob dies ein weiterer Schritt zum papierlosen Büro ist?

Anhang: Erkennungsverfahren

Feature Recognition

Unter Feature Recognition versteht man Verfahren, die versuchen, bei der Erkennung charakteristische Eigenschaften der Buchstaben zu entdecken. Ein "H" könnte z. B. als zwei senkrechte Linien verbunden mit einer horizontalen Querlinie beschrieben werden. Grenzen erreichen feature-basierte Technologien bei Buchstaben, zwischen denen auf Featurebasis kaum Unterschiede bestehen, wie z. B. der Buchstabe "O" und die Ziffer "0". Fast alle feature-basierten Systeme greifen daher zusätzlich auf Sprachtechnologie

oder Wörterbücher zurück und sind damit kontextabhängig. Feature-basierte Systeme sind meistens auf bestimmte Arten von Schriften (z. B. serifenlose Schriften) optimiert und zeigen erstaunliche Fehlerkennungen bei atypischen Schriften.

Pattern Matching

Beim Pattern Matching werden im System vorhandene ("gelernte") Muster auf Pixelebene mit neuen Mustern verglichen. Verschiedene Filterverfahren stellen sicher, daß nicht jedes vom Vergleichsmuster abweichende Pixel sofort als Abweichung gewertet wird.

Flexibilität durch freie Trainierbarkeit

Pattern Matching Systeme zeichnen sich durch hohe Flexibilität bei der Erkennung aus, da von seiten des Softwareherstellers kaum Vorgaben in Bezug auf die zu erkennenden Schriftarten vorgegeben werden. Optimale Ergebnisse erzielen diese Systeme allerdings immer erst nach einer mehr oder weniger aufwendigen Lernphase, weswegen sich die pattern-matching-basierten Verfahren überwiegend für Produktionseinsatz mit größeren Textmengen und hohen Anforderungen an die Erkennungssicherheit eignen.