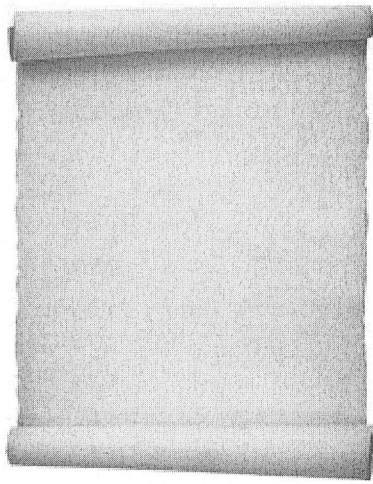


Frühere Schriftträger [Bearbeiten]

Höhlenzeichnungen sind die ältesten Dokumente, die der Mensch mit Pigmentfarbe auf einen Untergrund gezeichnet hat. Die Sumerer, die Träger der ältesten bekannten Hochkultur, schrieben seit etwa 3200 v. Chr. mit Keilschrift auf weiche Tontafeln, die zum Teil bis heute überliefert sind, weil sie durch Zufälle gebrannt wurden. Auch aus Ägypten sind Schriftträger aus anorganischen Materialien bekannt, beispielsweise die Prunkpalette des Königs Narmer (3100 v. Chr.) aus Speckstein.



10 Papier

Papyrus besteht aus den flach geschlagenen, über Kreuz gelegten und gepressten Stängeln der am gesamten unteren Nil in ruhigen Uferzonen wachsenden Schilfpflanzen (Echter Papyrus). Geschrieben wurde darauf mit schwarzer und roter Farbe. Die schwarze Tusche bestand aus Ruß und einer Lösung von Gummi arabicum, die rote Farbe wurde auf Ocker-Basis hergestellt. Das Schreibgerät

war ein Pinsel aus Binsen. Zwar gab es Papyrus im antiken Griechenland, jedoch war eine Verbreitung über Griechenland hinaus kaum bekannt. Im 3. Jahrhundert ersetzten die Griechen den Pinsel durch eine gespaltene Rohrfeder. Von dem griechischen Wort *pápyros* leitet sich das Wort Papier ab.

Titel

Im Römischen Reich wurde sowohl Papyrus benutzt als auch Wachstafeln. In die Letzteren wurde der Text mittels angespitzter Griffel geritzt. Nach dem Auslesen wurde das Wachs mit einem Schaber geglättet und die Tafel konnte erneut beschrieben werden. Öffentliche Verlautbarungen wurden meist als dauerhafte Inschrift an Tempeln oder Verwaltungsgebäuden angebracht. In Indien wurden Blätter von Palmengewächsen und in China Tafeln aus Knochen, Muscheln, Elfenbein benutzt. Später bestanden Schrifttafeln auch aus Bronze, Eisen, Gold, Silber, Zinn, Jade, Steinplatten und Ton oder auch aus organischem Material, wie Holz, Bambusstreifen und Seide. Pflanzenblätter und Tierhäute wurden noch nicht als Schriftträger benutzt. Orakelknochen wurden mit Griffeln geritzt oder mit Tinte mit Lampenruß oder Zinnober als Pigment beschriftet. Im mittelalterlichen Europa wurde dann vermehrt Pergament aus Tierhäuten benutzt. Die Tierhäute werden mit Pottasche oder Kalk gebeizt, gründlich gereinigt und aufgespannt getrocknet, es folgte das Schaben und die Oberflächenbearbeitung.

German

Texterkennung ist deshalb notwendig, weil optische Eingabegeräte (Scanner oder Digitalkameras, aber auch Faxempfänger) als Ergebnis ausschließlich Rastergrafiken liefern können, d. h. in Zeilen und Spalten angeordnete Flächen unterschiedlicher Färbung (Pixel). Texterkennung bezeichnet dabei die Aufgabe, zusammengehörende Farbflächen, die Buchstaben darstellen, als solche zu erkennen. Automatische Texterkennung und OCR werden im deutschen Sprachraum oft synonym verwendet. In technischer Hinsicht bezieht sich OCR jedoch nur auf den Teilbereich der Mustervergleiche von separierten Bildteilen als Kandidaten zur Erkennung von Einzelzeichen. Diesem OCR-Prozess geht eine globale Strukturerkennung voraus, in der zuerst Textblöcke von graphischen Elementen unterschieden, die Zeilenstrukturen erkannt und schließlich Einzelzeichen separiert werden. Diese OCR-Ergebnisse können durch nachfolgende Verfahren, in denen der Kontext berücksichtigt wird, präzisiert werden.

Ursprünglich wurden zur automatischen Texterkennung eigens entworfene Schriftarten entwickelt, die zum Beispiel für das Bedrucken von Scheckformularen verwendet wurden. Diese Schriftarten waren so gestaltet, dass die einzelnen Zeichen von einem OCR-Lesegerät schnell und ohne großen Rechenaufwand unterschieden werden konnten. So zeichnet sich die Schriftart OCR-A (DIN 66008, ISO 1073-1) durch einander besonders unähnliche Zeichen, besonders bei den Ziffern, aus. OCR-B (ISO 1073-2) ähnelt mehr einer serifenlosen, nicht-proportionalen Schriftart, während OCR-H (DIN 66225) handgeschriebenen Ziffern und Großbuchstaben nachempfunden wurde.

kopf

Die gestiegene Leistungsfähigkeit moderner Computer und verbesserte Algorithmen erlauben inzwischen auch die Erkennung von „normalen“ Druckerschriftarten bis hin zu Handschriften (etwa bei der Briefverteilung); wenn jedoch Lesbarkeit durch Menschen nicht vorrangig ist, werden drucktechnisch und erkenntnisstechnisch einfacher handhabbare Strichcodes genutzt.

Moderne Texterkennung umfasst inzwischen mehr als reine OCR, das heißt die Übersetzung einzelner Schriftzeichen. Zusätzlich werden Methoden der Kontextanalyse Intelligent Character Recognition (ICR) hinzugezogen, mit welcher die OCR-Ergebnisse korrigiert werden können. So kann eine „8“ zu einem „B“ korrigiert werden, wenn sie in einem Buchstaben-Kontext steht. Das OCR-Ergebnis „8aum“ kann mit ICR-Methoden zu „Baum“ korrigiert werden. Im Bereich industrieller Texterkennungssysteme wird daher von OCR/ICR-Systemen gesprochen. Die Grenzen des OCR-Begriffes sind jedoch fließend, denn OCR und ICR dienen auch als Marketingbegriffe, um technische Weiterentwicklungen besser vermarkten zu können. Auch Intelligent Word Recognition (IWR) fällt unter diese Kategorie. Dieser Ansatz versucht das Problem bei der Erkennung von Fließhandschriften zu lösen, bei der die Einzelzeichen nicht eindeutig separiert und daher nicht über herkömmliche OCR-Methoden erkannt werden können. Mit dieser Methode lassen sich jedoch keine Volltexte, sondern nur inhaltlich eng definierte und in ihrer Varianz bekannte Textausschnitte erkennen. Je größer die Varianz, desto zufallsabhängiger wird die Erkennungsgenauigkeit. Da IWR-Methoden nicht für Volltexterkennung geeignet sind, gilt generell, dass Fließhandschrift maschinell nicht lesbar ist. Das gilt auch für Schriften von Sprachen, die keine klaren Buchstabentrennungen haben, wie zum Beispiel die arabische Schrift.