

Literarische und Dokumentarische Datenverarbeitung

OPTOPUS Lesesystem / Scanner

Aus Sachmitteln des FSP 08 »Wissenschaftliche Textdatenverarbeitung« wurde das in BI 89/1 schon kurz vorgestellte Lesesystem OPTOPUS («**O**ptisch **o**rientierter **P**rozessor zur Umsetzung von **S**chriften») der Firma Makrolog beschafft und am 6. 12. 1989 in Raum 014 installiert.

Das System besteht aus einem IBM-kompatiblen PC mit dem Betriebssystem Concurrent DOS, einem Scanner HighScan 40 (400 dpi) der Firma SIEMENS, einem Koprozessor und der Lesesoftware.

Dank der größeren Flexibilität des OPTOPUS gegenüber dem nach wie vor vorhandenen KDEM sind damit die Einlesemöglichkeiten für gedruckt vorliegende Texte wesentlich erweitert, vor allem was Texte betrifft, die in Sonderschriften (kursiv, Fraktur) und in nicht-lateinischen Alphabeten vorliegen.

Diese Flexibilität war das entscheidende Auswahlkriterium. Sie wird dadurch erreicht, daß das System nicht mit voreingestellten festen Mustern arbeitet oder von solchen ausgeht, sondern grundsätzlich »trainiert«, d. h. auf die zu lesenden Schriften eingestellt werden muß. Trainingssätze können jedoch aufbewahrt und bei gleichartigen Vorlagen wiederverwendet werden.

OPTOPUS unterscheidet vier Arbeitsmodi:

Trainingsmodus: Dies ist der Lernmodus des Systems. In dieser Phase nimmt der Benutzer die Zuordnung von Zeichenmustern zu Buchstaben oder Buchstabenkombinationen und zu einem Zeichenfont vor. Jedes nicht erkannte (weil noch nicht trainierte) Zeichenmuster wird angezeigt; der Benutzer gibt die Bedeutung des angezeigten Zeichens ein, die zusammen mit dem Muster im Zeichenfont gespeichert wird.

Interaktiver Modus: In diesem Modus liest das System auf der Basis des aktuellen Trainingsstandes. Nicht oder nicht sicher erkannte Zeichen werden angezeigt und können nachtrainiert werden. Der interak-

tive Modus findet immer Anwendung, wenn ein weitestgehend fehlerfreies Leseergebnis erforderlich ist.

Vollautomatischer Modus: Die gesamte gescannte Seite wird ohne Rückfrage bearbeitet; nicht sicher erkannte Zeichen werden markiert und können nach dem Lesevorgang ergänzt bzw. korrigiert werden.

Modus »unkorrigiert«: Arbeitsweise wie im vollautomatischen Modus, jedoch werden eventuelle Fehlerstellen nicht markiert.

Verwendung als Scanner

Die mitgelieferte HighScan-Software der Firma Siemens, die unter Microsoft Windows 2.03 läuft, erlaubt ein Ablegen eingescannter Abbildungen zur Weiterverarbeitung in einer Desk-Top-Publishing-Umgebung. Die Auflösung ist zwischen 50 dpi und 400 dpi in 1 dpi-Schritten einstellbar. 64 Graustufen lassen sich unterscheiden. 99 Ausschnitte können definiert, verschoben, kopiert, gelöscht, vergrößert oder verkleinert werden. Inversdarstellung, Spiegelung, Rasterung von Grauwertbildern mit 6 verschiedenen Rastern ist möglich.

Die Weiterverarbeitung wird durch folgende Dateiformate unterstützt:

- TIFF (z. B. für PageMaker)
- RIFF (z. B. für ImageStudio)
- EPSF (z. B. für Adobe Illustrator)
- PCX (z. B. für PC-Paint)
- MSP (z. B. für MS-Paint)
- IMG (z. B. für GEM, Ventura Publisher)
- STS (z. B. für Stylesheets).

Für die Benutzung des Lesesystems ist eine Terminvereinbarung notwendig. Interessenten wenden sich bitte werktags in der Zeit zwischen 9 und 11 Uhr an Herrn Ziegler (Tel. 3405).

Sobald mehr Erfahrung im Umgang mit dem System vorliegt, ist eine Informationsveranstaltung zum Einsatz dieses Systems geplant. Der Termin wird in der nächsten BI angekündigt.