

Literarische und Dokumentarische Datenverarbeitung

TUSTEP unter UNIX auf PCs

TUSTEP steht jetzt auch unter Interactive UNIX für 386er und 486er PCs zur Verfügung. Diese Version umfaßt wie die übrigen UNIX-Versionen den vollen Leistungsumfang von TUSTEP.

Leistungsfähige PCs sind damit in der TUSTEP-Umgebung eine Alternative zu RISC-Workstations. Unsere Messungen haben ergeben, daß TUSTEP unter Interactive UNIX auf dem selben Rechner etwa doppelt so schnell läuft wie unter DOS. Ein mit 50 MHz getakteter

486er Rechner mit 16 MB Hauptspeicher und 1 GB Festplatte erreicht unter TUSTEP etwa die Leistung einer IBM RS6000 Modell 310 oder einer Silicon Graphics Indigo und etwa 25% mehr als eine Sun SPARC 2.

Interessant kann eine solche Lösung u. a. in den Fällen sein, in denen eine DOS-kompatible Tastaturbelegung eine Rolle spielt. Die Tastaturfunktionen unter DOS und unter Interactive UNIX sind in TUSTEP identisch.

Neues Texterkennungs-System (Scanner + Software)

Neben dem Texterkennungssystem OPTOPUS, das in BI 89/3 vorgestellt wurde, ist seit einigen Wochen ein weiteres Texterkennungssystem im Einsatz. Es besteht aus einem 386er PC mit einem Scanner HP ScanJet IIc und der Texterkennungssoftware OmniPage Professional.

Dieses System soll das weiterhin verfügbare OPTOPUS ergänzen, insbesondere für das Erkennen von Texten, die ausschließlich in lateinischen Zeichen (einschließlich Umlaute und Akzentbuchstaben) gedruckt sind. Es kommt im Gegensatz zum OPTOPUS für diesen Einsatzbereich in der Regel ohne Training aus, kann aber für Nicht-Standard-Zeichen (z.B. mathematische Symbole) zusätzlich trainiert werden.

Im Gegensatz zu OPTOPUS arbeitet OmniPage Professional nicht mit dem Vergleich von vorgefundenen und trainierten Zeichen-Formen (pattern matching), sondern mit der Analyse von Eigenschaften der vorgefundenen Zeichen (feature recognition).

Dies hat u. a. den bereits erwähnten Vorteil, daß auf die beim OPTOPUS unabdingbare Trainingsphase verzichtet werden kann. Wei-

terer Vorteil: Das Erlernen der Bedienung dieses Systems ist sehr viel einfacher.

Nachteil: Die Unterscheidung verschiedener Schriftgrößen ist mit diesem Programm ebenso wenig möglich wie die Unterscheidung verschiedener Fonts (lediglich Normalschrift, Kursivschrift und fette Schrift werden auseinandergelassen). Für viele Anwendungen kann auf diese Unterscheidungen jedoch verzichtet werden.

Für »Standard«-Anwendungen steht hiermit ein Lesesystem zur Verfügung, das den Einsatz der automatischen Texterkennung auch für kleinere Textmengen lohnend macht. Ein automatischer Papiereinzug erleichtert zusätzlich die Bedienung, wenn der Text auf Einzelblättern (z. B. Fotokopien) vorliegt.

Die mit dem HP ScanJet selbst mitgelieferte Software DeskScan-II erlaubt die Bearbeitung von Bildinformationen wie: Bilder vergrößern oder verkleinern, drehen, spiegeln, mehrere Bilder kombinieren, aber auch: Bildschärfe verbessern, Grauwerte anpassen und (bei Farbbildern) Farben ändern.