

Literarische und Dokumentarische Datenverarbeitung

Seit 1970 hat die Abteilung LDDV als Aufgabe »die Entwicklung und Betreuung von Methoden und Programmen zur Verarbeitung von Textdaten aller Art. Dadurch soll vor allem den geisteswissenschaftlichen Fachbereichen der Universität Zugang zu dem für sie relativ neuen Hilfsmittel EDV ermöglicht werden...«.

Diese Arbeiten werden seit März 1985 vom Ministerium für Wissenschaft und Kunst Baden-Württemberg im Rahmen des Forschungsschwerpunkts 08 »Wissenschaft-

liche Textdaten-Verarbeitung« zusätzlich gefördert.

Die »Kommission Forschung Baden-Württemberg 2000« hat in ihrem Abschlußbericht vom Juli 1989 eine Weiterförderung empfohlen. Begründung: »Tübingen hat auf dem Gebiet der Textverarbeitung, das zunehmend an Wichtigkeit für alle Geisteswissenschaften gewinnt, eine führende Stellung. Sie sollte der Universität unbedingt erhalten bleiben.« (S. 196).

TUSTEP

Mit TUSTEP, dem »Tüebinger System von Textverarbeitungs-Programmen«, versucht die Abteilung LDDV den oben genannten Aufgaben Rechnung zu tragen. TUSTEP enthält über die heute übliche »Textverarbeitung« (Eingeben, Korrigieren, Suchen, Formatieren) hinaus Funktionen, die es zu einem Werkzeug für die Wissenschaften machen, in denen Texte Objekt der wissenschaftlichen Arbeit sind: für Philologien, Sprachwissenschaften, Literaturwissenschaften, historische Wissenschaften, Bibliothekswesen; Wissenschaften also, in denen nicht nur neue Texte als Produkt der eigenen wissenschaftlichen Arbeit erstellt und publiziert werden sollen, sondern in denen schon existierende, überlieferte, schriftlich fixierte oder zu fixierende Texte (einschließlich literarischer Texte und historischer Quellen) durch kritische Neuedition gesichert, sprachlich und stilistisch analysiert, inhaltlich erschlossen, bibliographisch erfaßt werden müssen.

Dem tragen Grundoperationen der Textdaten-Verarbeitung (und entsprechende TUSTEP-Programme) Rechnung, die mit folgenden Schlagworten grob charakterisiert werden können: *Zeige-* und *Such-*Funktionen (im Editor und in den übrigen Programmen) mit leistungsfähiger Muster-Erkennung auch für linguistische Fragestellungen; *Vergleichen* von verschiedenen Textfassungen; *Korrigieren* nicht nur interaktiv im Editor, sondern auch anhand vorbereiteter (u.U. automatisch erstellter) Korrekturanweisungen; *Zerlegen* von Texten in (vom Benutzer zu definierende) Elemente (z.B. Wortformen); *Sortieren* von Textelementen oder von längeren Texteinheiten nach einer

Vielzahl von Alphabeten und anderer Kriterien; *Register erstellen* durch Zusammenfassen sortierter Textelemente; *Bearbeiten* von Textdaten durch vom Benutzer definierte Regeln zum Auswählen, Ersetzen, Umstellen, Ergänzen, Zusammenfassen, Vergleichen von Textteilen, durch Rechnen mit Zahlenwerten, die bereits im Text enthalten sind (z.B. Kalenderdaten) oder aus ihm gewonnen werden können (z.B. die Zahl der Wörter in einem Satz), und Ausgeben in verschiedenen Formaten, auch solchen, die von anderen Systemen (z.B. SPSS zur statistischen Auswertung) benötigt werden.

Den genannten Aufgaben tragen auch die erlaubten Dateilängen und die Zahl der möglichen Angaben (z.B. von Suchbegriffen) Rechnung; sie liegen in Bereichen, die normalerweise keine praktische Bedeutung haben dürften (z.B. maximale Größe einer Datei: 8000 MByte).

Aufgaben, die mit TUSTEP bearbeitet werden, reichen vom Schreiben einer Seminararbeit bis hin zum Erstellen von umfangreichen Bibliographien, Lexika, Indizes, Konkordanzen, Wörterbüchern, Editionen und natürlich auch Monographien, jeweils einschließlich der automatischen Herstellung der Druckvorlagen für diese Werke in der vom Buchdruck gewohnten Qualität (z. Zt. über DIGISET oder Monotype LASERCOMP; Laserdrucker-Ausgabe wird für Publikationen in der Regel nur zum Korrekturlesen und zum Überprüfen des späteren Satzbildes zu empfehlen sein).

Neben den Programmen für die Grundoperationen der Textdaten-Verarbeitung enthält TUSTEP auch eine Reihe organisatorischer Leistungen, wie sie üblicherweise

vom Betriebssystem eines Rechners bereitgestellt werden. Dies ermöglicht es, alle für die Textdaten-Verarbeitung notwendigen Funktionen, einschließlich der Datenhaltung und Datensicherung, auf Rechnern mit unterschiedlichen Betriebssystemen in gleicher Weise aufrufen zu können und erspart damit dem Benutzer beim Wechsel auf einen Rechner mit einem anderen Betriebssystem nicht nur ein Umlernen, sondern erlaubt

auch, bereits erstellte TUSTEP-Kommandofolgen unverändert zu übernehmen.

TUSTEP läuft derzeit unter den Betriebssystemen MVS (IBM), VM/CMS (IBM), VMS (DEC VAX), MS-DOS (IBM-PC's und kompatible); die Implementierung unter BS2000 (SIEMENS) und UNIX ist geplant. TUSTEP steht inzwischen außer in Tübingen auch an etwa 50 anderen Hochschulen zur Verfügung.

Neues von TUSTEP

Seit dem 18. 9. 1989 gibt es eine erweiterte (596 S.) und verbesserte Neuauflage des TUSTEP-Handbuchs. Sie ist zum Preis von 16,- DM bei Frau Krier im ZDV erhältlich.

Was ist neu im Handbuch?

- Die OS1100-Version (UNIVAC) ist eingefroren und nicht mehr beschrieben.
- Die Besonderheiten der VMS- und der MS-DOS-Version, die gegenüber der MVS- und VM-Version vor allem im Editor wesentliche Verbesserungen beim Eingeben und Korrigieren durch bessere Cursor-Steuerung enthält, sind beschrieben.
- Die von TUSTEP standardmäßig unterstützten Zeichensätze sind erweitert. Sie

umfassen jetzt neben dem lateinischen, griechischen, hebräischen und altsyrischen Alphabet und mathematischen Sonderzeichen auch das kyrillische Alphabet sowie den phonetischen Zeichensatz (IPA).

- Einige Kommandos wurden in ihren Leistungen wesentlich erweitert (u.a. der Editor durch »Datenbank-Anweisungen« für ein Abfragen von Datenbeständen in datenbank-ähnlicher Art); neue Kommandos sind hinzugekommen (u.a. das Programm EINFUEGE, mit dem Textteile, die in einer Datei stehen und über Kürzel identifiziert werden, in die Daten einer anderen Datei eingefügt werden können).

Aktuelles

Optopus-Testinstallation: Bis Mitte November steht in den Räumen des FSP 08 ein Zeichenerkennungssystem OPTOPUS der Firma Makrolog zur Erprobung bereit. Es besteht aus einem Scanner HighScan (400 dpi) der Fa. Siemens, einem AT-kompatiblen PC mit Koprozessor und der Lesesoftware. Das System muß grundsätzlich »trainiert«, d.h. auf die zu lesenden Zeichen eingestellt werden. Trainingssätze können jedoch auf-

bewahrt und bei gleichartigen Vorlagen wiederverwendet werden. - Das System erwies sich bei ersten Tests gegenüber dem seit Ende 1985 im FSP betriebene KDEM als sehr viel flexibler; u.a. konnten gedruckte Devanagari-Texte erfolgreich trainiert und gelesen werden. - Interessenten wenden sich bitte wegen Terminvereinbarung an Frau Straub (Tel. 6973 bzw. 3405).