

Portabilität von Programmen und Daten: eine nicht unwichtige Eigenschaft von TUSTEP

1. Datenverarbeitung in den Geisteswissenschaften

Als vor nunmehr 25 Jahren, am 1. September 1970, die Universität Tübingen die EDV-Unterstützung für die geisteswissenschaftlichen Fächer durch Einrichtung einer eigenen Abteilung mit der Aufgabe der »Entwicklung und Betreuung von Methoden und Programmen zur Verarbeitung von Textdaten aller Art« am Zentrum für Datenverarbeitung institutionalisiert hat, galten – nicht nur an deutschen wissenschaftlichen Rechenzentren – Anwender mit »nicht-numerischen« Problemen eher als Exoten.

Zwar gab es, lange bevor die meisten Universitäten überhaupt mit eigenen zentralen Rechenzentren ausgestattet waren, am Deutschen Rechenzentrum in Darmstadt (1961 als Stiftung der Bundesrepublik, des Landes Hessen und der DFG gegründet, um »allen Hochschulen und hochschulfreien Forschungsinstituten . . . die Bearbeitung wissenschaftlicher Probleme auf einer großen Rechanlage« zu ermöglichen) eine eigene Abteilung Nichtnumerik, untergliedert in die beiden Gruppen »Geisteswissenschaften« und »Informationswissenschaften«; zu seinen wissenschaftlichen Mitarbeitern zählten zwei Ägyptologen und ein Historiker. Dennoch sahen die meisten Rechenzentren damals ihre Aufgaben ausschließlich im eigentlichen »Rechnen«. »Textverarbeitung . . . auf den Rechanlagen z. Zt. nicht erlaubt«: so konnte man noch 1979 (als es diesen Begriff immerhin schon gab) im Jahresbericht des Rechenzentrums einer deutschen Universität lesen.

Auf diesem Hintergrund möge man die Weitsicht erlauben, die die Universität Tübingen vor 25 Jahren bewiesen hat, indem sie sich auch die Unterstützung der Wissenschaften, die nicht mit Zahlen, sondern mit Texten umgehen, explizit in den Aufgabenkatalog der Zentralen Einrichtung »Zentrum für Datenverarbeitung« geschrieben hat.

Schon zweieinhalb Jahre vorher hatte die Universität ihr »Rechenzentrum« offiziell in »Zentrum für Datenverarbeitung« umbenannt, um dem vollen Aufgabenspektrum Rechnung zu tragen. (Die Weichen waren bereits weitere zwei Jahre vorher gestellt worden, als 1966 das Rechenzentrum aus dem Mathematischen Institut ausgegliedert wurde und im gleichen Jahr

eine Stelle geschaffen wurde, auf der auch die Interessen der geisteswissenschaftlichen Fächer vertreten werden konnten.)

2. Das Tübinger System von Textverarbeitungsprogrammen

EDV-Unterstützung für Geisteswissenschaftler am ZDV ist heute für viele BI-Leser synonym mit TUSTEP. Dies ist zwar eine sehr einseitige Sicht, die viele Dienstleistungen des ZDV für diesen Nutzerkreis ausklammert. Richtig daran ist aber, daß das »Tübinger System von Textverarbeitungsprogrammen« aus der Beratung und Betreuung – und anfänglich auch Programmierung; fertige Software für den Umgang mit Textdaten gab es ja praktisch nicht – für Projekte aus den unterschiedlichsten geisteswissenschaftlichen Fachgebieten hervorgegangen ist.

Die wachsende Zahl und die Verschiedenartigkeit der nach Computer-Unterstützung suchenden Projekte machte es notwendig, vorgefertigte Bausteine bereitzustellen, aus denen der Anwender ohne Programmierkenntnisse selbst eine Lösung seines Problems zusammenstellen konnte. Um die erforderliche Flexibilität zu gewährleisten, mußte der Leistungsumfang der einzelnen Bausteine auf Grundfunktionen beschränkt werden. Gleichzeitig mußte es möglich sein, die Leistung dieser Bausteine an die jeweils gegebene Aufgabenstellung (z. B. die Sortierfolge an die Sprache des untersuchten Textes) anzupassen.

Im Herbst 1977 war es dann so weit: die Ausbildung der geisteswissenschaftlichen Computer-Anwender am ZDV konnte von der Unterrichtung im Gebrauch herkömmlicher Programmiersprachen und eines Unterprogramm Pakets zur Zeichen- und Stringverarbeitung in FORTRAN umgestellt werden auf Kurse im Einsatz dieser Bausteine. Seit 1978 tragen sie den Namen TUSTEP. Die Förderung, die diese Aktivitäten des ZDV im Rahmen des Forschungsförderungsprogramms des Landes Baden-Württemberg von 1985 bis 1989 erfahren haben, ermöglichte neben der inhaltlichen Weiterentwicklung eine Portierung auch auf andere Rechner, über die zentralen Server des ZDV hinaus; sie schuf die Voraussetzungen für die (und verpflichtete gleichzeitig zur) Weitergabe in größerem Umfang an andere Hochschulen und Forschungseinrichtungen.

3. Portabilität von Programmen und Daten

Wenn in den BIs des ZDV an dieser Stelle über TUSTEP berichtet wird, handelt es sich in der Regel um Hinweise auf neue Leistungen vorhandener Bausteine oder auf neue Bausteine, die aufgrund erweiterter Aufgabenstellungen hinzugekommen sind. Diese Funktionserweiterungen sind letztlich auch das, was die Anwender des Paketes interessiert. Vieles andere wird als selbstverständlich vorausgesetzt und hier nur bei aktuellen Anlässen (etwa bei einem Wechsel der zentralen Server) zur Sprache gebracht.

Zu diesen von TUSTEP-Nutzern eigentlich schon immer als selbstverständlich angesehenen Eigenschaften gehört die im Titel genannte Portabilität von Daten und Programmen. Der knappe Rückblick auf 25 Jahre Geschichte von TUSTEP soll Anlaß sein, ein paar Worte darüber zu verlieren.

Die Leistung, die ein Programmpaket für einen bestimmten Zweck erbringt, reicht alleine nicht aus, daß es erfolgreich für ein größeres Projekt eingesetzt werden kann. Ist ein Werkzeug nicht portabel, so kann der Wechsel eines bisher benutzten zentralen Rechners oder der Ersatz eines nicht mehr ausreichenden Arbeitsplatzrechners durch einen »modernerer« mit einem anderen Betriebssystem (manchmal reicht schon eine nicht wirklich aufwärtskompatible neue Version) empfindliche Einschnitte in der Arbeit noch nicht abgeschlossener Projekte bedeuten oder deren Fortführung ganz in Frage stellen. Die für ein Projekt erarbeiteten Prozeduren und die zugehörigen Daten problemlos von einem Rechner auf einen anderen portieren zu können ist Voraussetzung für die Kontinuität EDV-gestützter Arbeit. Gerade geisteswissenschaftliche Projekte sind besonders darauf angewiesen: sie sind einerseits häufig sehr langfristig angelegt; andererseits sind die dabei mit EDV zu bewältigenden Aufgaben in der Regel so komplex, daß es nicht ausreicht, ähnliche Basisfunktionen auch in der neuen Umgebung wieder vorzufinden; es muß vielmehr möglich sein, ganze Abläufe übertragen zu können.

TUSTEP bietet die dazu notwendigen Voraussetzungen:

- Aufwärtskompatibilität bei Leistungserweiterungen
- identische Benutzeroberfläche auf Rechnern verschiedener Architektur
- Einschluß der für den Programmablauf wichtigen organisatorischen Funktionen, die üblicherweise vom Betriebssystem bereitgestellt werden, vor allem zu Dateiverwal-

lung und Datensicherung und zur Ablaufsteuerung, in die TUSTEP-Oberfläche

- einheitliches Datenformat auf externen Datenträgern (Magnetband, Kassette, CD-ROM) für die verschiedenen Rechnerarchitekturen; Export-Format für den Austausch über Netze.

Es ist inzwischen fast sprichwörtlich, daß aufgrund dieser Eigenschaften TUSTEP-Nutzer in der Vergangenheit bei den Rechner-Wechseln im ZDV jeweils am wenigsten Probleme mit der Portierung ihrer Programme und Daten hatten. Selbst vom jeweils neuen Betriebssystem mußten sie kaum mehr kennen als die Prozeduren zum Einloggen und Ausloggen sowie den TUSTEP-Aufruf.

Die Frage der Portabilität gewinnt in der internationalen Diskussion der »Textwissenschaften« in den letzten Jahren immer mehr Bedeutung. Die Forderung nach Mehrfach-Verwendung von Dokumenten (z. B. zur konventionellen Publikation, zur Bereitstellung für die Absuche in Datenbanken, als Quelle sprachwissenschaftlicher Forschung) hat 1986 zur Standardisierung der Standard Generalized Markup Language SGML (ISO 8879) geführt. Eines der erklärten Ziele dieses Standards ist, daß nach dessen Vorschriften ausgezeichnete Dokumente nicht nur problemlos von einem Rechner auf einen anderen portiert werden können, sondern auch unabhängig von der für den ursprünglichen Zweck verwendeten Software sind. Seit 1987 arbeitet die Text Encoding Initiative (TEI), ein gemeinsames Unternehmen der Association for Computers and the Humanities, der Association for Literary and Linguistic Computing und der Association for Computational Linguistics, an Richtlinien für die Auszeichnung von Textdaten, die ein Standard-Format für den Datenaustausch in der geisteswissenschaftlichen Forschung darstellen sollen. 1994 wurden die »Guidelines for Electronic Text Encoding and Interchange« (Herausgeber: Michael Sperberg-McQueen und Lou Burnard) veröffentlicht.

TUSTEP hat schon immer die Mittel bereitgestellt, die die Verwendbarkeit von Texten zu mehr als einem Zweck garantieren. Die in der Literatur (seit Textverarbeitung auf Computern eine Rolle spielt) häufig beklagte Trennung von Textverarbeitung (was meist identisch war mit Formatierung oder Desk Top Publishing) und Datenverarbeitung (meist in Form von Datenbank-Anwendungen) gab es in TUSTEP nicht.

Es hängt freilich vom Anwender ab, wie weit er für die Mehrfach-Verwendung seiner Texte Sorge trägt. Hinsichtlich der Konventionen, nach denen er seine Texte inhaltlich gliedert und

mit entsprechenden Markierungen versieht, macht TUSTEP keinerlei Vorschriften. Die Fachgebiete und die Arbeitsziele, mit denen dort Texte erfaßt, analysiert und verarbeitet werden, sind zu unterschiedlich; der Aufwand, für andere als das zunächst vorgesehene Arbeitsziel Markierungen im Text vorzusehen, ist gleichzeitig für viele (oft personell sehr knapp ausgestattete) Projekte zu hoch. Deswegen ist TUSTEP völlig offen, was Art und Umfang der Auszeichnung (»markup«, »encoding«) angeht. Und deswegen stellen auch nach SGML- oder TEI-Konventionen ausgezeichnete Texte kein Problem dar, weder was die Analyse und Verarbeitung solcher Texte angeht (vgl. BI 95/1+2 S. 13 bezüglich der typographischen Aufbereitung von Dokumenten, die nach TEI-Richtlinien ausgezeichnet sind) noch was die Umwandlung eines für ein Projekt gewählten und ausreichend detaillierten Kodierungs-Schemas in eine SGML-konforme Auszeichnung betrifft.

4. »Making a Synthesizer Sound like an Oboe«

Dies ist der Untertitel des Hauptvortrags, den Michael Sperberg-McQueen, PhD, am 18.11.1995 beim 65. Kolloquium zur Anwendung der EDV in den Geisteswissenschaften an der Universität Tübingen halten wird und zu

dem hiermit herzlich eingeladen wird. Anlässlich des Rückblicks auf 25 Jahre EDV-Unterstützung der Geisteswissenschaften am ZDV soll sein Vortrag »The State of Computing in the Humanities« den Blick über den Tübinger Horizont hinaus lenken. Die im vorliegenden Beitrag angesprochene Thematik wird dabei nicht zu kurz kommen: Sperberg-McQueen ist als Geisteswissenschaftler (Germanist) am Rechenzentrum der University of Illinois in Chicago tätig und einer der Väter der TEI-Guidelines: neben Lou Burnard von der Universität Oxford und einem internationalen steering committee war er maßgeblich für deren Erarbeitung, Redaktion und Publikation verantwortlich.

Unmittelbar vor dem Kolloquium, vom 15.–17. November, wird Sperberg-McQueen zusammen mit Winfried Bader vom ZDV einen Workshop »SGML-konforme Textauszeichnung nach den Richtlinien der Text Encoding Initiative (TEI) veranstalten (siehe die ausführliche Ankündigung weiter hinten in diesem Heft). Dort werden die Voraussetzungen für hardware- und software-unabhängige Portabilität der Daten einen zentralen Platz einnehmen.

Wilhelm Ott
ott@zdv.uni-tuebingen.de