

Der IBM-705 war eines der stärksten Datenverarbeitungssysteme, die Mitte der 1950er Jahre auf den Markt kamen. Er gehörte zur IBM 700/7000 Großrechner-Bauserie, die der Konzern zwischen 1952 und den 1960er Jahren anbot.



Rückblick

Die Anfänge der Digital Humanities

Von hartnäckigen Vorurteilen, Theologen als Pionieren und einer späten Dynamik: Bis die Geisteswissenschaftler vom Wert digitaler Ressourcen überzeugt waren, dauerte es Jahrzehnte.

VON KURT GÄRTNER

DIE DIGITALEN Geisteswissenschaften, auch *Humanities Computing* oder – wie man sie seit etwa zehn Jahren auch im deutschen Sprachraum überwiegend bezeichnet – *Digital Humanities* (DH), erleben in jüngster Zeit einen enormen Aufschwung. Tagungen, Kolloquien, Workshops und Publikationen, gedruckt wie digital, gibt es in Fülle, und fast jede Woche werden neue Stellen in den geisteswissenschaftlichen Fächern ausgeschrieben, die mit dem Attribut „digital“ eine Art Wende in der Forschungspraxis manifestieren. Blickt man zurück auf die Anfänge dieser Entwicklung, dann ist es erstaunlich, wie lange es gedauert hat, bis es zur allgemeinen Akzeptanz des neuen Werkzeugs und zugleich auch des Mediums, des Computers, in den Geisteswissenschaften kam. Im folgenden Rückblick berichte ich hauptsächlich über Erfahrungen aus den Philo-



Pionier der EDV: der Jesuit Roberto Busa (links).
Rechts: IBM-Gründer Thomas J. Watson.



logien, insbesondere auch aus meiner eigenen Disziplin, der Deutschen Philologie.

Theologen als Pioniere

Die Bedeutung des Computers für den Einsatz in den Geisteswissenschaften wurde bereits in den 1950er Jahren erkannt. Zu den Pionieren der Anwendung der Elektronischen Datenverarbeitung, der EDV, wie es damals hieß, gehörte der Jesuit Roberto Busa (1913–2011). Mit Unterstützung des IBM-Gründers Thomas J. Watson (1874–1956) gelang es ihm zwischen 1950 und 1966, das Gesamtwerk des Thomas von Aquin maschinenlesbar zu machen. Die Texte wurden auf Lochkarten eingegeben und über einen Lochkartenleser im Computer gespeichert. In

einem halbautomatischen Verfahren wurden sie dann lemmatisiert, d. h. die flektierten Wortformen wurden auf ihre Grundform, das Lemma, zurückgeführt bzw. ihm zugeordnet. Für diesen Arbeitsschritt hatte Busa ein elektronisches Wörterbuch entwickelt, das Lexicon Electronicum Latinum. Das von 1974 bis 1980 durchgeführte Unternehmen gipfelte in dem sukzessiv publizierten monumentalen Index Thomisticus, der 49 Foliobände umfasste und im modernen elektronischen Lichtsatzverfahren gedruckt war. Dazu kamen noch sieben weitere Bände mit den Referenztexten, den echten Werken des Thomas sowie den ihm

Standardlochkarte mit
80 Spalten.

zugeschriebenen Texten. Diese Pionierleistung bedeutete einen enormen Fortschritt für die Reputation der EDV in den Geisteswissenschaften. Die Theologen, denen wissenschaftliches Edieren und die lexikographische Erschließung des Bibeltextes von Haus aus vertraut sind, haben nicht nur die Anfänge, sondern auch die weitere Entwicklung der Digitalen Geisteswissenschaften maßgeblich mitgeprägt.

Busas Lexicon Electronicum Latinum wurde auch zur Lemmatisierung bei der Erstellung der 1976 publizierten Vulgata-Konkordanz durch den Benediktiner Bonifatius Fischer (1915–1997) benutzt, einem Vorhaben, das er in Kooperation mit Wilhelm Ott vom Zentrum für Datenverarbeitung der Universität Tübingen durchführte. Die Textbasis der Konkordanz war die neue Stuttgarter Ausgabe der Vulgata, die unter maßgebender Mitarbeit von Fischer entstanden war. Der Text der Ausgabe wurde wie in Busas Projekt über Lochkarten maschinenlesbar gemacht: Um einen möglichst fehlerfreien Text zu erhalten, wurde er jedoch zweimal eingegeben, anschließend automatisch verglichen und das entstandene Fehlerprotokoll manuell abgearbeitet. Bei diesem und den bis

Programmsteuerkarte, auf einer rotierenden Trommel aufgespannt zur Beschleunigung der Datenerfassung.

hin zur Publikation erforderlichen weiteren



Arbeitsschritten des Vorhabens wurde das im Tübinger Zentrum entwickelte Programmpaket TUSTEP benutzt. Das Verfahren der Doppeleingabe hat sich später bei der Retrodigitalisierung großer Wörterbücher wie z. B. des 33 Bände umfassenden Deutschen Wörterbuchs der Brüder Grimm bewährt. Der dadurch gewonnene maschinenlesbare Text übertrifft an Genauigkeit in jeder Hinsicht das Ergebnis einer nur einmaligen Erfassung, auch wenn diese mehrfach Korrektur gelesen wurde.

Als ein von Philologen für Philologen entwickeltes Werkzeug war TUSTEP besonders attraktiv für Editoren und Lexikographen, die mit älteren Texten und ihrer großen Zeichenvielfalt zu tun hatten. Durch ein einfaches und leicht zu beherrschendes Kodierungsverfahren ließen sich so gut wie alle Zeichen, auch solche mit Sub- und Superscripten, wie sie häufig z. B. in mittel-



hochdeutschen Texten auftreten, darstellen und korrekt alphabetisieren.

Digitales Textarchiv in Cambridge

Neben Roberto Busa gehörte der englische Germanist Roy Wisbey zu den großen Pionieren, die die Anfänge der Digitalen Geisteswissenschaften maßgebend bestimmt haben. Wie Busa erkannte auch Wisbey früh, welche Möglichkeiten der Computer für die lexikographische Arbeit bot. Voraussetzung für seinen Einsatz waren maschinenlesbare Texte. Wisbey gründete 1960 das Centre for Literary and Linguistic Computing an der Universität Cambridge und etablierte dort ein digitales Textarchiv, das hauptsächlich mittelhochdeutsche Texte umfasste. Diese bildeten die Basis für die Generierung von zahlreichen Konkordanzen und Indices. Sämtliche in den 1960er und 1970er Jahren in Cambridge zunächst über Lochkarten, dann über Lochstreifen maschinenlesbar gemachten Texte sind noch heute

erhalten und über verschiedene Institutionen frei zugänglich.

Korrekte maschinenlesbare Texte auf der Basis der besten kritischen Ausgaben waren für die EDV-Pioniere eine unabdingbare Voraussetzung für alle weitergehenden literarischen und linguistischen Analysen. Die damals geschaffenen und bis heute verfügbaren digitalen Texte wurden immer wieder gebraucht und daher auch gepflegt, deshalb haben sie den raschen technologischen Fortschritt ohne Qualitätsverlust überstanden. Digitale Ressourcen, die genutzt werden, bleiben verfügbar.

Anfängliche Mängel ...

Der frühe Einsatz des Computers für lexikographische Zwecke hatte zur Folge, dass die Index- und Konkordanzherstellung bald einen ungeahnten Aufschwung nahm. Während zwischen 1938 und 1964 nur rund ein Dutzend manuell erstellter Indices und eine Konkordanz zu deutschen Texten erschienen, sind von 1966,

Lochkartenlocher 029 von IBM mit Bedienstation.

nach der Publikation des ersten automatisch erzeugten Index, bis 1996 neben drei manuell hergestellten rund 90 mit EDV produzierte Indices und Konkordanzen von teilweise beträchtlichem Umfang und Gewicht, allerdings auch unterschiedlicher Machart und Qualität, herausgekommen.

Besonders eklatant waren die Mängel in der Alphabetisierung, weil der Standardzeichenvorrat für die Computerausdrucke zunächst auf Großbuchstaben beschränkt war und/oder in den Sortieralphabeten der Computerhersteller z. B. die deutschen Umlaute ä ö ü und das ß nicht berücksichtigt waren. Man behalf sich in der Regel damit, *ä ö ü ß* in *AE OE UE SS* oder *SZ* aufzulösen, was natürlich zu Verzerrungen und Abweichungen im Hinblick auf die in der deutschen Lexikographie übliche Alphabetisierung führte. Dabei waren die Sortierprobleme schon Ende der 1960er Jahre durch Roy Wisbey und Wilhelm Ott gemeistert worden.

Ein weiterer Mangel der frühen Indices und Konkordanzen war u. a. die Beschränkung ihrer Makrostruktur auf Wortformen, d. h. sie waren nicht lemmatisiert, was vor allem wegen der morpho-

logischen Struktur der deutschen Verbalflexion ein Problem war. Auch die Homographentrennung war nicht vollautomatisch durchführbar; die Unterscheidung der verschiedenen Bedeutungen einer Wortform blieb letzten Endes einem linguistisch kompetenten Bearbeiter überlassen. Nur wenige der älteren Indices und Konkordanzen sind wie der Index Thomisticus und die Vulgata-Konkordanz lemmatisiert. Seit 2000 werden auch lemmatisierte Indices und Konkordanzen nur noch ausnahmsweise in Buchform publiziert, denn für alle lexikographischen Hilfsmittel hat sich das Internet als ideales Medium erwiesen, zumal es die Verknüpfung mit anderen digitalen Lexika ermöglicht.

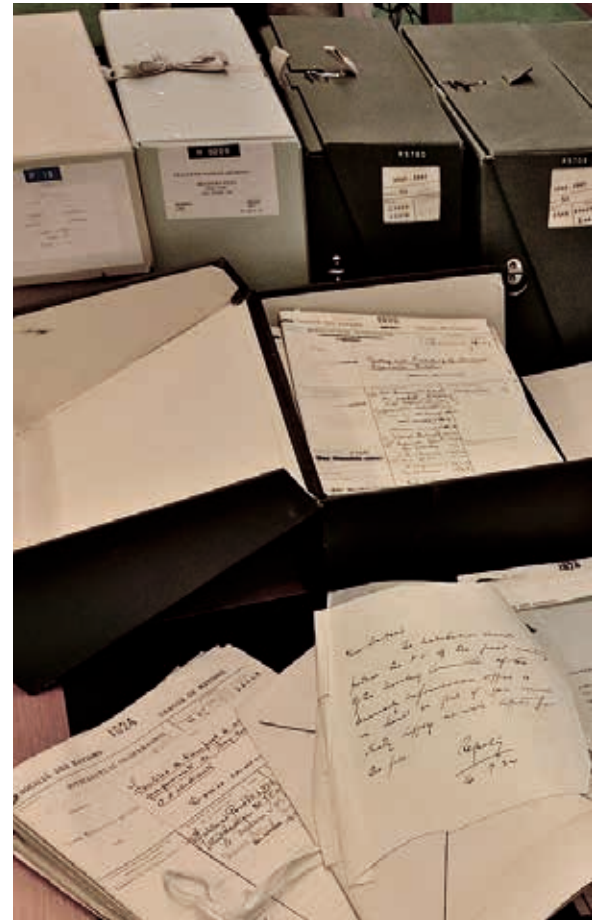
... und große Vorteile

Früh erkannte man den Wert von Indices und Konkordanzen auch für die Erstellung von Editionen. Denn sie verzeichneten mit einer vorher kaum erreichbaren Vollständigkeit und Genauigkeit die Wortschatzbelege mit zugehörigen Stellenangaben und Verteilungen. Daraus ergab sich die Möglichkeit, jedes beliebige Wort und jede Wortform mit Sicherheit wieder aufzufinden. Anhand eines Wortformenindex konnte man z. B. die Normalisierung eines Textes überprüfen und Fehler korrigieren. Diesen Vorteil sahen auch Herausgeber, die vom Computer als Editionswerkzeug partout nichts wissen wollten. War die Arbeit an der Konstituierung des kritischen Textes abgeschlossen und dieser maschinenlesbar gemacht, konnte man mit Hilfe eines digitalen Wortformenindex Inkonsistenzen und letzte Fehler im edierten Text leicht auffinden und ausbügeln.

Von der Brauchbarkeit des Computers für solche speziellen Zwecke waren also sogar traditionelle Herausgeber leicht zu überzeugen, auch wenn sie sonst den Möglichkeiten des neuen Werkzeugs für die Kollation und sprachliche Analyse der Textzeugen noch nichts abgewinnen konnten. Man wollte verständlicherweise weiterarbeiten wie immer. Ebenso wenig interessierte die Technik, die für die Herstellung von Konkordanzen und bald auch für den Druck einer Edition zur Verfügung stand.

Hartnäckige Vorurteile

Auch wenn die neue Technologie für die Forschungspraxis attraktive Perspektiven eröffnete, waren bis zum Ende der 1990er Jahre die Vorurteile gegen den Einsatz des Computers in der Textphilologie immer noch weit verbreitet. Das trug wohl mit dazu bei, dass in den Editionswissenschaften und bei der Erstellung



der großen historisch-kritischen Ausgaben, wie sie im Akademienprogramm seit je gefördert wurden, wesentliche Entwicklungen „verschlafen“ wurden und dass das neue Werkzeug in vielen Projekten erst verspätet zum Einsatz kam. Zu den ersten vorurteilslosen Anwendern der neuen Technologie gehörten ausgerechnet sogenannte „Altgermanisten“. Während meiner Zeit am Londoner King's College (1972–1974) verfassten Roy Wisbey und ich einen Beitrag mit dem Titel „Die Bedeutung des Computers für die Edition altdeutscher Texte“. Der Beitrag erschien 1974 und lag ganz im Trend einer Serie von fünf zwischen 1971 und 1997 durchgeführten internationalen Kolloquien zur maschinellen Verarbeitung altdeutscher Texte. Die Einrichtung computerlesbarer Textarchive war in den 1970er Jahren bereits eine Selbstverständlichkeit, ebenso ihr unbestreitbarer Nutzen für die Lexikographie. Die Debatte um den Einsatz des Computers in der Editionswissenschaft aber zog sich länger hin. Es gab wohl in der amerikanischen Presse Schlagzeilen wie „Editing at the Speed of Light“

DER AUTOR

Prof. Dr. Kurt Gärtner ist Mitherausgeber des neuen Mittelhochdeutschen Wörterbuchs, das an den Akademien in Göttingen und Mainz erstellt wird. Bis 2010 war er Projektleiter des Vorhabens „Deutsche Texte des Mittelalters“ der Berlin-Brandenburgischen Akademie der Wissenschaften und bis 2015 Vorsitzender sowie bis 2015 Mitglied der Arbeitsgruppe „Elektronisches Publizieren“ der Union der deutschen Akademien der Wissenschaften, der Vertreter aller Akademien angehören.



und übertriebene Erwartungen unter denen, die von der Sache nichts verstanden. Unser Beitrag zielte jedoch darauf ab, die damals nur von Wenigen mit genügender Schärfe erkannten Vorteile automatischer Verfahren für die Edition mittelalterlicher deutscher Texte in ein helleres Licht zu rücken und für einen umfassenden Einsatz des Computers als Werkzeug und Medium auf allen Stufen der Editionsarbeit zu plädieren.

Einzug des PCs in die Textwissenschaften

Das Misstrauen der Textphilologen gegenüber der neuen Technologie wurde erst allmählich aufgebrochen, als ab Mitte der 1980er Jahre der *Personal Computer*, der PC, seinen Einzug in die Sekretariate hielt, schließlich auch auf die Schreibtische der Gelehrten gelangte und – auch wenn einige von ihnen das Ding nicht einmal einzuschalten wussten – doch zumindest den Anschein der Akzeptanz des Neuen signalisierte. Die wenigen Geisteswissenschaftler, die seit den 1960er Jahren mit der neuen Technologie und ihrer stürmischen Entwicklung vertraut waren, führten ihre Computerarbeiten auf Großrechnern, den *mainframes*, durch und waren in der Lage, große Textmen-

gen zu verarbeiten. Dazu war der PC, bestückt mit den kommerziellen Textverarbeitungsprogrammen wie WORD oder WordPerfect, anfangs nicht in der Lage, aber wegen der Korrekturmöglichkeiten, die er eröffnete, war er in den Sekretariaten von enormem Vorteil, weil das Neuabschreiben von Texten obsolet geworden war. Für die wissenschaftliche Textdatenverarbeitung wurde dann mit der raschen Entwicklung leistungsfähiger Geräte und Software auch der PC interessant, insbesondere für die Erfassung der Daten, die für die Weiterverarbeitung in geeignetere Formate wie XML konvertiert werden konnten.

Auch um die Jahrtausendwende musste jedem editorisch tätigen Germanisten klar sein, dass seine nur im Druckmedium publizierte Edition über eine digitale Vorstufe erzeugt wurde. Die Fixierung auf die Druckfassung führte allerdings auch noch im digital aufgeklärteren 21. Jahrhundert dazu, dass die Satzdaten nach dem Erscheinen einer Edition wegge-

worfen wurden, ob mit Absicht, weil man um die Zukunft des Buches fürchtete, oder ob aus Unwissen um die damit verbundenen Potentiale, bleibe dahingestellt. Das ist heute wohl nicht mehr so recht vorstellbar. Doch wie mein Rückblick zeigt, dauerte es Jahrzehnte, bis man vom Wert digitaler Ressourcen überzeugt war. ■

Inzwischen überall angekommen: PC und Notebook als Werkzeug und Medium.

Literatur und WWW

R. Busa: The Annals of Humanities Computing: The Index Thomisticus. In: Computers and the Humanities 14 (1980), 83–90. www.alice.id.tue.nl/references/busa-1980.pdf

R. Wisbey: Computer und Philologie in Vergangenheit, Gegenwart und Zukunft. In: Maschinelle Verarbeitung altdeutscher Texte IV. Beiträge zum Vierten Internationalen Symposium, Trier 28. Februar bis 2. März 1988, hrsg. von K. Gärtner, P. Sappeler und M. Trauth, Tübingen 1991, 346–361.

S. Hockey: Electronic Texts in the Humanities – Principles and Practice, Oxford 2000.