

Bewertungskriterien für elektronische Editionen

von Fotis Jannidis

(Alle Rechte an diesem Text liegen beim Autor.)

- [1. Allgemeines](#)
- [2. Das elektronisch erfaßte Textkorpus](#)
- [3. Die elektronische Erschließung des Textes](#)
 - [a. Textkodierungen, Textauszeichnung](#)
 - [b. Information Retrieval](#)
 - [c. Navigation](#)
- [4. Das Programm](#)
- [5. Weiterführendes](#)

Die Beurteilung gedruckter Texte und Editionen hat eine lange Tradition in der Literaturwissenschaft, und die entsprechenden Kriterien sind – bis auf einige umstrittene Punkte – mehr oder weniger selbstverständlicher Teil des Fachs. Ganz anders sieht die Lage für die digitalen Medien aus: Elektronische Editionen sind neue Gegenstände im philologischen Alltag, und ihre Gestalt scheint zur Zeit noch eben so schwankend wie die Bewertungskriterien der Anwender solcher Editionen. Andererseits steigt seit wenigen Jahren die Menge der verfügbaren CD-ROMs, es liegen also inzwischen mehrere Beispiele wissenschaftlich brauchbarer Texte vor, und – was vielleicht noch wichtiger ist – sie werden auch verwendet. Insbesondere die Arbeit an der Online-Zeitschrift ›[Computerphilologie](#)‹, die regelmäßig CD-ROM-Rezensionen bringt, verdeutlichte den Bedarf einer Klärung, welche Aspekte bei der Besprechung digitaler Editionen zu berücksichtigen seien. Im Nachfolgenden wird ein solcher Kriterienkatalog diskursiv entwickelt – in der Hoffnung, andere finden ihn interessant genug, um ihn weiterzudiskutieren. Im Anhang findet sich eine Stichwortliste mit den genannten Bewertungsaspekten.

1. Allgemeines

Vorweggeschickt sei, daß alle Standards der Bewertung gedruckter Texte für elektronische Editionen ebenfalls gelten – auf einige Probleme, die dadurch entstehen, wird noch ausführlicher einzugehen sein –, aber nur wenige davon im nachfolgenden erwähnt werden, nämlich nur dann, wenn sie für elektronische Editionen besonders oder anders relevant sind. So ist etwa vom angezielten Adressaten und vorgeschlagenen Verwendungszweck der Editionen nicht nur die vorauszusetzende Textqualität abhängig, sondern auch die Langfristigkeit der Edition oder der Reichtum der Recherchemöglichkeiten.

Elektronische Texte erscheinen nicht nur alleinstehend, sei es als CD-ROM oder im Internet, sondern öfter auch zusammen mit einem Buch, in dem das Ganze oder Teile des elektronischen Texts abgedruckt sind. Solche Hybrideditionen strukturieren ihr Material je nach den Verwendungszwecken, die im jeweiligen Medium realisiert sind (gängigerweise die intensive Lektüre im Drucktext, das Textretrieval oder die mehrsichtige Darstellung komplexer Textzusammenhänge im elektronischen Text).

2. Das elektronisch erfaßte Textkorpus

Elektronische Editionen können alle Beziehungen zu Vorlagen aufweisen, die auch gedruckte haben können. Sie können sich also auf Handschriften, Typoskripte, frühere Drucke usw. beziehen; sie können eine vollständige Erfassung aller einschlägigen Texte oder eine Auswahl mit entsprechenden Kriterien darstellen. Sie können im Vergleich mit den Vorlagen normalisiert und modernisiert sein und andere editorische Eingriffe aufweisen. Wenn es sich um Retrodigitalisierungen gedruckter Editionen handelt, sind sie darüber hinaus aber von einer bestimmten modernen Druckvorlage abhängig. In diesem Fall kann, wenn die Vorlagen für die gedruckte Ausgabe nicht noch einmal konsultiert wurden, die Textqualität stets nur schlechter sein als im Druckexemplar. Digitalisierung ist nämlich fehlerfrei nicht zu haben; die Frage ist nur: wieviele Fehler sind geblieben? Die Antwort, die auch vom gewählten Digitalisierungsverfahren abhängig ist, kann nicht so einfach gegeben werden, da Zufallsfunde nichts über die tatsächliche Fehlerdichte in der Edition aussagen. Das Ideal der Fehlerfreiheit verstellt hier etwas den Blick auf die notwendigen Abstriche, die auch sehr zeitaufwendige Druckeditionen von historisch-kritischen Ausgaben machen müssen und die für weniger kostenintensive Ausgaben entsprechend anzusetzen sind.

Oftmals wird in Besprechungen von elektronischen Editionen hervorgehoben, ob und inwieweit diese die Multimedialität des neuen Mediums ausreizen. Diese Herangehensweise tendiert manchmal dazu, die neuen Freiheiten zu einem neuen Zwang zu machen, d.h. eben deshalb, weil viele Medien im neuen Medium verbunden werden können, muß dieses Potential auch ausgenutzt werden. Die Diskussion darüber ist oftmals auch verbunden mit der Feststellung, daß elektronische Texte zur längeren Lektüre nicht taugen, die Digitalisierung von Texten daher also nur in einer entsprechenden Multimedialisierung ihren Sinn finden könne. Der nun nahezu zehnjährige Umgang mit solchen Multimedialprodukten hat allerdings deutlich gemacht, daß diese nur für einen Anwendungsbereich wirklich geeignet sind: für die Wissensvermittlung. Wissenschaftliche Editionen, die Werkzeug zur Wissenserzeugung sein sollen, können von solchem Aufpolieren kaum gewinnen, da die Filme, Bilder und Töne meist nur illustrativen und exemplarischen Zweck haben, und nicht selbst wiederum, z.B. in der Form einer Bilddatenbank, Recherchemittel sind. Die oben genannten typischen Verwendungsweisen, die sich inzwischen für den wissenschaftlichen Umgang mit elektronischen Texten etabliert haben, sind davon unberührt. Kurzum, es gibt wohl keinen Grund, von den Eigenschaften des Mediums auf notwendige Eigenschaften der Editionen zu schließen, vielmehr sind diese vom Verwendungszweck abhängig zu machen. Die Wiedergabe von Ton, Film und Bild innerhalb elektronischer Editionen kann also auch ganz unterschiedlichen Zwecken dienen, wovon zumeist ihre Auswahl und auch die Qualität ihrer Wiedergabe abhängig ist.

Eine elektronische Edition, wie auch jede gedruckte, erschließt einen Text und macht ihn zugänglicher, aber immer um den Preis, daß dabei auch Informationen verloren gehen. Das kann technische Gründe haben, wie z.B. die Wiedergabe nichtlateinischer Alphabete und komplexerer handschriftlicher Befunde, da diese in älteren Zeichensätzen und mit den üblichen Darstellungsmitteln nur annäherungsweise repräsentiert werden können. Das hat aber auch systematische Ursachen: Was wird digitalisiert, welche Aspekte der Vorlage werden in der Digitalisierung repräsentiert (typischerweise Strukturinformationen, die den Text untergliedern, z.B. Werk, Kapitel), welche werden als unwesentlich beiseite gelassen (typischerweise Angaben zu Zeilenumbruch und -zählung, Papierqualität, typographische Gestaltung der Seite usw.) und wird diese Differenz reflektiert?

3. Die elektronische Erschließung des Textes

Die elektronische Erschließung des Textes ist insbesondere von drei Faktoren abhängig: a. Von der Textauszeichnung, b. von den Möglichkeiten zum Information Retrieval, die die verwendete Software bietet, und c. von der Navigationsstruktur

a) Textkodierungen, Textauszeichnung

Elektronische Texte bestehen stets aus einem Text, den der angezielte Leser auch lesen können soll, und Zusatzinformationen zu diesem Text. Zwei Beispiele zur Verdeutlichung: Der Satz "Das ist sehr wichtig!" enthält eine Reihe von Buchstaben, nämlich "Das ist sehr wichtig!", und eine zusätzliche typographische Information: die Unterstreichung von "sehr", die hier wohl eine Betonung wiedergibt. Wenn auf einem Buchumschlag die Zeichenkette: "Johann Wolfgang Goethe: Die Leiden des jungen Werthers" zu finden ist, dann versteht das ein Leser heutzutage als Angabe von Autor und Titel des im Buch abgedruckten Werks. Solche zusätzlichen Informationen sind also zum einen Teil des Gesamttexts, müssen aber vom Lesetext unterschieden werden. Nur wenn diese zusätzlichen Informationen in den Text eingetragen sind, der Text also ausgezeichnet wurde, kann man in einer elektronischen Edition darauf zugreifen. Das gilt für so triviale Fälle wie die Unterstreichung, die nur dann auch auf dem Bildschirm dargestellt werden kann, wie auch für anspruchsvollere Aspekte, z.B. die Angaben zu Autor und Titel, nach denen man nur so in einem größeren Korpus suchen kann. Die Textauszeichnung ist für den ›Leser‹ einer Ausgabe, der diese ja fast immer mittels eines Erschließungsprogramms wahrnimmt, zumeist nur in Form von typographischen Merkmalen und Such- bzw. Navigationsmöglichkeiten merkbar. Eine Textauszeichnung, die nur von einem Programm gelesen werden kann, wird jedoch dazu führen, daß der Text unbrauchbar geworden ist, sobald dieses Programm aufgrund von Betriebssystem- oder anderen Softwareentwicklungen nicht mehr lauffähig ist. Das Verfahren, mit dem die Textauszeichnung notiert wird, bestimmt also darüber, wie lange und wie problemlos man auf solche Texte zugreifen kann. Man unterscheidet daher zwischen einer proprietären Kodierung, deren Verfahren evtl. sogar unbekannt sind und die den Text eng an ein Programm koppelt, und einer offenen Kodierung, die sich auf einen offenen, dokumentierten und zugänglichen Standard stützt.

Wie schon erwähnt, bestimmt die Textauszeichnung darüber, welche Informationen zusätzlich zum reinen Textfluß in einer elektronischen Edition überhaupt optisch, im Textretrieval oder für die Navigation wahrgenommen werden können. Sie bestimmt damit also ganz wesentlich die Verwendungsmöglichkeiten des Textes. Jede digitale Edition muß schon aus pragmatischen Gründen eine Auswahl treffen, welche Aspekte der Vorlage kodiert werden sollen und welche nicht; sie muß Entscheidungen darüber treffen, inwieweit sie den Text mittels Textauszeichnung für bestimmte Fragestellungen aufbereitet, z.B. für ein effizienteres Information Retrieval durch die Eintragung von Grundformen von Wörtern und der Auflösung von Mehrdeutigkeiten (z.B.: Er ging <Grundform="gehen"> zur Bank <Oberbegriff="Finanzinstitut">.) Die Textauszeichnung legt über den Text meist auch ein oder mehrere hierarchische Netze, auf die zur Navigation und für das Information Retrieval zurückgegriffen wird: z.B.: ›Autor - Gattung - Einzelwerk - Kapitel/Akt‹ oder ›Buch - Seite - Zeile‹. Diese Strukturierungen lehnen sich oftmals an herkömmliche Verfahren an, wie sie aus dem Bereich des Drucks bekannt sind, allerdings erfordern die besonderen Möglichkeiten und Probleme digitaler Editionen oft auch neue Lösungen, und insgesamt steht die einmal gewählte Lösung, da sie stets nur eine von mehreren Möglichkeiten ist, unter einem ganz anderen Legitimationsdruck.

b) Information Retrieval

Das Information Retrieval, also die Informationsgewinnung durch Suchverfahren, ist sicherlich die häufigste Verwendungsweise digitaler Editionen. Ein wesentliches Beurteilungskriterium für eine digitale Edition liegt daher in der Qualität des Information Retrieval. Sehr elaboriert sind inzwischen die Möglichkeiten, nach Zeichenketten zu suchen, sehr viel weniger entwickelt dagegen die Verfahren zur Suche nach Bedeutungseinheiten.

Die üblichen Verfahren des Information Retrieval reichen weit über die einfache Suche nach Zeichenketten hinaus, wie sie den meisten aus der Textverarbeitung bekannt ist. Weit verbreitet sind Worttrunkierungen, also die Suche nach Bruchstücken von Worten mittels eines Sonderzeichens, das entweder für beliebig viele beliebige Zeichen oder für ein einziges beliebiges Zeichen steht. (Bsp.: die Suche nach "geh*" findet die Worte "gehen", "geht", "geheim", wenn das Sternchen als 'Joker' für beliebig viele Zeichen steht.) Ebenso geläufig ist inzwischen, wohl aufgrund der Internet-Suchmaschinen, die Verwendung der Booleschen Operatoren UND, ODER und NICHT. Eine Textstelle qualifiziert sich erst dann als Treffer, wenn die so formulierte Bedingung erfüllt ist. (Bsp.: "Herz UND Schmerz" findet alle Stellen, in denen beide Worte vorkommen.)

Sehr viel seltener sind Suchmaschinen zu finden, die mit regulären Ausdrücken umgehen können – einer Technik, Klassen von Zeichenketten zu finden, die eine Erweiterung der Verwendung von Jokern darstellt. Ebenso rar sind Retrievalprogramme, die einen Thesaurus verwenden, so daß bei einer Suche nach "Frau" auch alle Textstellen gefunden werden, in denen "Weib" vorkommt. Da Literaturwissenschaftler häufig mit historischen Texten arbeiten, ist ein Thesaurus, der die verschiedenen möglichen Schreibweisen eines Wortes ("Sein" und "Seyn", "Tür" und "Thür") bei einer Abfrage verwendet, schon eine große Hilfe. Das gleiche gilt für Programme, die automatisch disambiguieren und den Wortstamm ermitteln können.

Ein weiterer Aspekt eines Retrievalprogramms ist die Möglichkeit, den Skopus der Suche zu bestimmen, also ihn entweder einzuengen, z.B. nur in den Werken eines Autors, nur in einem bestimmten Werk oder nur in den Sprechtexten einer Dramenfigur zu suchen, oder auszuweiten, indem andere Korpora mit dem vorliegenden verbunden werden können. Hierzu gehört die Frage, ob die Suchmaschine den freien Zugriff auf die Textauszeichnungen erlaubt oder nur einige ausgewählte für die Suche verwendet werden können. Text Retrieval Programme, die für philologische Zwecke erstellt wurden, erlauben oftmals eine statistische Auswertung der Suchergebnisse, stellen also nicht nur die absoluten Zahlen der Treffermenge dar, sondern bieten darüber hinaus Informationen, die die Ergebnisse im Kontext des ausgewerteten Textes sehen.

Das Erstellen einer komplexen Abfrage ist, schon weil es für die hier genannten Verfahren kaum Standards gibt, meist eine mühselige Angelegenheit, die aber über gute Hilfsfunktionen, digitale Assistenten und andere gute Geister erleichtert werden kann. Manche Programme bieten auch die Möglichkeit, umfangreichere Abfragen abzuspeichern, um sie wieder- und weiterzuverwenden.

Die Präsentation und Weiterverarbeitung der Suchergebnisse ist für die alltägliche Verwendung wohl fast ebenso wichtig, wie die Suche selbst. Sieht man auf einen Blick, wo im Gesamttext die Treffer sind und wieviele es sind? Kann man schnell von Trefferstelle zu Trefferstelle springen? Kann man die Liste durch weitere Abfragen, nur über die

Treffermenge, verkleinern? Kann man weitere statistische Prozeduren anschließen? Nicht zuletzt: Kann man die Trefferstellen abspeichern?

c) Navigation

Das Bewegen im Text mittels Inhaltsverzeichnissen, Hyperlinks und Navigationsleisten stützt sich auf die ausgezeichneten Textelemente, die von der Software verwendet werden. Oben wurde bereits das Problem angesprochen, daß jede Ordnungsstruktur andere Möglichkeiten der Ordnung verdrängt. Anders als im Fall von gedruckten Editionen kann man elektronische Editionen jedoch so gestalten, daß auf dasselbe Datenmaterial unterschiedliche Sichtweisen möglich sind und damit auch zwischen verschiedenen hierarchischen Ordnungsstrukturen hin- und hergeschaltet werden kann. Für die netzartigen Hypertextstrukturen stellt sich dagegen eher das Problem, ob diese für den Anwender plausibel ausgewählt und eingerichtet sind, d.h. ob die Kriterien der Vernetzung auch nachvollziehbar sind und tatsächlich Ordnung erzeugt wird und nicht Unübersichtlichkeit. Hypertexte haben sich insgesamt nämlich als schwer überschaubar erwiesen, da die üblichen Orientierungsmarkierungen zur momentanen Position im Text (Finger zwischen den Seiten usw.) fehlen. Solche Markierungen können die Programme auf ganz unterschiedliche Weise im unkörperlichen Raum des elektronischen Textes setzen und teilweise auch bei Bedarf ausblenden.

4. Das Programm

Wenn die elektronische Edition gleich mit einem Erschließungsprogramm geliefert wird, dann weist dies neben denen der Suchmaschine noch eine Reihe von anderen Eigenschaften auf, die für den ›Leser‹ von Bedeutung sind.

Für die Weiterverarbeitung ist der Export von Textdaten entweder direkt aus dem Korpus oder aus der Treffermenge eines Suchlaufs notwendig. Über die Brauchbarkeit einer solchen Exportfunktion entscheiden eine ganze Reihe von Faktoren: In welchem Umfang ist er möglich? Welche Dateiformate werden unterstützt? Können also Formatierungen u.ä. direkt in die Textverarbeitung übernommen werden? Kann die Textauszeichnung mitexportiert werden, etwa zur Weiterverarbeitung des Textes in anderen Retrievalprogrammen oder zur statistischen Auswertung? Insbesondere bei teureren Editionen, die nur in Bibliotheken installiert werden, sind Funktionen wichtig, die es erlauben, alle Treffer einer Suchabfrage mit frei definierbarer Kontextgröße und einschließlich der genauen Angabe ihrer jeweiligen Position im Text auf einmal abzuspeichern, damit der Anwender sie in Ruhe zu Hause bearbeiten kann.

Insbesondere sehr umfangreiche Editionen können sich langfristig zu einem richtigen Arbeitswerkzeug für den Anwender entwickeln, insbesondere wenn sie ihm erlauben, in dem elektronischen Text zu arbeiten wie in einem Buch. Dazu gehört die Möglichkeit, Textstellen anzustreichen oder Anmerkungen zu machen. Solche Features sind manchmal verbunden mit den elektronischen Möglichkeiten, dann kann man alle Anstreichungen suchen oder nur den Text der Anstreichungen durchsuchen lassen. Da auch in diesem Fall das Problem des schnellen Verfalls elektronischer Editonen besteht, muß der Benutzer seine Anmerkungen auch einschließlich der zugehörigen Textstelle exportieren können.

Elektronische Editionen sind wie ihre gedruckten Gegenstücke nicht fehlerfrei, allerdings konservieren sie nicht wie diese ihre Fehler, sondern können optimiert werden. Korrekturen

können durch Textupdates etwa übers Internet verbreitet werden. Vielleicht ist dem Anwender sogar selbst der Eingriff in den Text möglich. Das Gesagte gilt ebenso für die verwendete Software, die zum einen fast immer Fehler enthält und zum anderen verbesserungsfähig ist. Elektronische Editionen sind also mit ihrer Publikation nicht abgeschlossen, sondern im Prinzip offene Projekte, die durch die Rückmeldung der Anwender in Text und Programm verbessert werden können.

Insgesamt gehört zu den üblichen Bewertungskriterien von Software auch deren Benutzerfreundlichkeit; dazu zählt neben der schon angesprochenen ergonomischen Gestaltung des Textes und der Unterstützung bei der Formulierung von Suchausdrücken auch die Programmoberfläche, die Verfügbarkeit eines brauchbaren Handbuchs und evtl. eine Adresse bei Problemen mit der Edition. Nicht unerwähnt bleiben dürfen natürlich auffallende Besonderheiten, z.B. besondere Hardwarevoraussetzungen oder ein Kopierschutz und Abweichungen von den üblichen Erwartungen, etwa besonders schnelle oder langsame Geschwindigkeit des Programms oder des Information Retrievals.

Weiterführendes

Kurt Gärtner / Peter Kühn: Indices und Konkordanzen zu historischen Texten des Deutschen: Bestandaufnahmen, Typen, Herstellungsprobleme, Benutzungsmöglichkeiten. In: Werner Besch / Anne Betten / Oskar Reichmann / Stefan Sonderegger (Hg.): Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. Berlin, New York: de Gruyter 21998, S. 715-742.

[Besonders wichtig der Abschnitt zu den Herstellungsproblemen]

F.J.: Goethes Werke auf CDROM. Chadwyck-Healey 1995. In: Arbitrium 16,2 (1998), S. 192-201.

[Rezension eines elektronischen Texts auf der Grundlage einer früheren Fassung der vorliegenden Überlegungen]

Waltraud Maierhofer: Konkordanzen und Elektronische Bibliothek als Hilfsmittel der Goetheforschung. In: Wirkendes Wort 42 (1992), S. 420-29.

Eric Johnson: Oxford Electronic Text Library Edition of the Complete Works of Jane Austen. In: Computers and the Humanities 28 (1984), S. 317-321.

[Rezension eines elektronischen Texts, der ohne Programm vertrieben wird]

Diskussion von elektronischen Texten:

[Humanist](#)

[e-edition](#)

Rezensionen zu elektronischen Texten:

[Computers and the Humanities](#)

[Computerphilologie](#)

[H-Soz-u-Kult](#)

Text online verfügbar bei:

[Sic et Non. Zeitschrift für Philosophie und Kultur.](#)