

Elektronische Editionen

Möglichkeiten, Grenzen, Perspektiven

INAUGURALDISSERTATION

zur Erlangung des Grades eines Doktors der Philosophie
dem Fachbereich Philosophie und Geisteswissenschaften
der Freien Universität Berlin

vorgelegt von

Martin Mertens

aus Düsseldorf

Jahr der Einreichung: 2008

1. Gutachter: Prof. Dr. Ursula Kocher

2. Gutachter: Prof. Dr. Wolfgang Neuber

Tag der Promotion: 11.07.2008

Inhalt

Vorbemerkung	3
Einführung	5
1 Elektronische Editionen	9
1.1 Typen elektronischer Editionen	9
1.2 Entstehungsbedingungen elektronischer Editionen	12
1.3 Buchnahe elektronische Editionen	14
1.3.1 Datenerfassung	16
1.3.2 Datenverarbeitung	16
1.3.3 Datenausgabe	16
1.4 Elektronischen Medien nahe stehende elektronische Editionen	17
1.4.1 Navigation	18
1.4.2 Textdarstellung	20
1.4.3 Hypertext und Hypermedia	22
1.4.4 Retrieval	23
1.4.5 Textexport	28
1.4.6 Notizen, Lesezeichen und Exzerptdatenbanken	29
1.4.7 Retrodigitalisierung	29
1.5 Hybrideditionen	31
1.6 »Status quo« elektronischer Editionen	32
1.6.1 Verfügbarkeit elektronischer Editionen	32
1.6.2 Tabellarischer Überblick	33
2 Text in Büchern und elektronischen Medien	37
2.1 Texte in Büchern.	38
2.1.1 Geschichte als Heuristik	39
2.1.2 Text nimmt materielle Gestalt an	41
2.1.3 Text und Gestaltung	43
2.1.4 Buchproduktion und Distribution	55
2.1.5 Entmaterialisierung und Virtualisierung von Texten	58
2.2 Texte in elektronischen Medien	58
2.2.1 Speicherung von Texten in elektronischen Medien	59
2.2.2 Datenformat und Darstellungsformat	59
2.2.3 Kodierung elektronischer Texte	60
2.2.4 Datenformate für formatierte Texte	60
2.2.5 Semantische Kodierung	63
2.2.6 X-Media-Publishing	64
2.3 Elektronische Editionen und Editionswissenschaft.	65
2.3.1 Einrichtung vorhandener Editionen für elektronische Medien	66
2.3.2 Erarbeitung zukunftssicherer elektronischer Editionen	67
2.3.3 Produktionsseitige Dokumentation	68

3	Konzepte und Technologien für elektronische Editionen	69
3.1	Auszeichnungssprachen	71
3.1.1	Grundprinzip von Auszeichnungssprachen	71
3.1.2	Standard Generalized Markup Language	74
3.1.3	Hyper Text Markup Language	76
3.1.4	Extensible Markup Language	77
3.1.5	XML-basierte Datenformate	78
3.1.6	Auszeichnungssprachen in den Geisteswissenschaften	79
3.2	Extensible Markup Language	80
3.2.1	Einführung in XML	81
3.2.2	XML Grundbegriffe	81
3.2.3	Darstellungsweisen	86
3.2.4	XML-Inhaltsmodelle	87
3.2.5	XML-Namensräume	88
3.2.6	XML-Anwendungen	91
3.2.7	XML-Technologien	92
3.3	Die Text Encoding Initiative.	94
3.3.1	TEI-Richtlinien	95
3.3.2	TEI-Inhaltsmodell	96
3.3.3	Textkodierung nach TEI-Richtlinien	97
3.3.4	Beispiel aus Franz Kafka: DER PROCESS	97
3.3.5	Beispiel aus Georg Büchner: DANTONS TOD	108
3.4	Metadaten	126
3.4.1	Erhebung von Metadaten	127
3.4.2	Speicherung von Metadaten	129
3.4.3	Verwendung von Metadaten	129
3.4.4	Kodierung von Metadaten	131
4	Ausblick	137
	Glossar	139
	Literatur	143
	Internetressourcen	156

Vorbemerkung

Diese Arbeit stellt Konzepte und Technologien für elektronische Editionen auf dem Gebiet der Germanistik vor. Der Schwerpunkt der Arbeit liegt darauf darzustellen, welche Informationen schriftliche Texte enthalten, die über den buchstäblichen Informationsgehalt hinausgehen und wie diese Informationen für elektronische Editionen kodiert und elektronischen Auswertungsmöglichkeiten zugänglich gemacht werden können. Die vorgestellten Konzepte zur Kodierung von Texten für elektronische Editionen beruhen auf dem Konzept von *Auszeichnungssprachen*, im Wesentlichen der *Extensible Markup Language (XML)* in Verbindung mit den Richtlinien der *Text Encoding Initiative (TEI)*.

Neben philologischen und editionswissenschaftlichen Aspekten macht die Beschäftigung mit Computern und Computertechnologien einen nicht unerheblichen Teil dieser Arbeit aus. Kenntnisse auf dem Gebiet der Computertechnik werden jedoch nicht vorausgesetzt. Wo technische Grundlagen zum Verständnis der Arbeit notwendig sind, werden diese in die Darstellung einbezogen. Häufig benutzte Begriffe und Abkürzungen aus dem Gebiet der EDV sind zusätzlich in einem Glossar am Ende des Textes erläutert. Begriffe, die im Glossar nachgeschlagen werden können, werden im Text kursiv dargestellt.

Die Darstellung beschränkt sich auf Möglichkeiten zur Kodierung bzw. Auszeichnung von Texten, um sie mit Hilfe geeigneter Computerprogramme für die philologische Arbeit nutzen zu können. Die Entwicklung von Programmcode oder eigenständiger Programme wird nicht thematisiert – sie gehört in das Gebiet der Informatik. Die Auszeichnung von Texten darf dagegen als ureigenstes Gebiet der Textedition gelten. Anstelle typografischer Mittel kommen im Kontext elektronischer Editionen lediglich andere Verfahren der Textauszeichnung, die *Tags* einer *Auszeichnungssprache*, zum Einsatz.

Die Beispiele in dieser Arbeit sind im Internet unter <http://www.e-ditionen.de/diss/examples.html> abrufbar.

Einführung

Computer und elektronische Medien bestimmen längst den Arbeitsalltag der meisten Germanisten. Texte werden auf dem Computer geschrieben, Exzerpte in Datenbanken gesammelt, ein großer Teil der Korrespondenz wird per E-Mail erledigt.

Die Analogien zwischen Computerprogrammen und Schreibmaschine, Zettelkasten oder Brief zeigen, dass Computer im täglichen Umgang hauptsächlich als komfortabler Ersatz für Aufzeichnungen auf Papier benutzt werden. Es ist leichter, Texte im Textverarbeitungsprogramm zu ändern und zu korrigieren, als auf Papier. Eine Datenbank nimmt weniger Platz ein und lässt sich schneller durchsuchen als ein Zettelkasten oder gar Schränke voller Karteikarten. Ein elektronischer Brief landet nahezu ohne Zeitverzögerung auf dem »Desktop«, dem elektronischen Schreibtisch des Empfängers.

Wenn es aber um Texte als Gegenstände der wissenschaftlichen Arbeit geht, wird der wissenschaftliche Diskurs noch weitgehend durch Bücher geprägt. Überspitzt betrachtet erscheint die Situation paradox: Texte werden mit Hilfe von Computern erstellt und auf Papier publiziert. Dabei werden die Texte bis hin zur Druckmaschine in Form elektronischer Daten verarbeitet. Werden die gedruckten Texte zur wissenschaftlichen Arbeit herangezogen, werden sie wieder vom Papier in den Computer überführt. Einzelne Textstellen werden als Exzerpte gespeichert und als Zitate in eigenen Texten verwendet; ganze literarische Werke werden elektronisch erfasst, um mit Hilfe von Computern Wortindizes und Konkordanzen zu erstellen. Die Ergebnisse der Arbeit mit dem Computer werden wiederum auf Papier publiziert.

Diese Darstellung ist grob vereinfacht. Der größte Teil des Materials, das für die wissenschaftliche Arbeit herangezogen wird, angefangen bei Quellen, Textzeugen und Textausgaben bis hin zum weitaus größten Teil der wissenschaftlichen Literatur, liegt in Form von Papier vor und wurde nicht mit Hilfe von Computern bearbeitet. Angesichts des zunehmenden Einsatzes elektronischer Medien in den Geisteswissenschaften stellt sich aber die Frage, warum elektronische Textsammlungen bisher nur eine geringe Bedeutung erlangt haben.

Die Frage, warum elektronische Editionen in der Germanistik kaum Beachtung finden, war der Anlass, diese Arbeit zu schreiben. Die oben geschilderte Situation bei der wissenschaftlichen Arbeit lässt offen, warum Texte so selten elektronisch publiziert werden, wenn sowohl die Erstellung von Texten als auch die Arbeit mit den Texten weitgehend mit elektronischen Hilfsmitteln erfolgt. Dazu sollen im Folgenden Möglichkeiten, Grenzen und Perspektiven elektronischer Editionen untersucht und aufgezeigt werden.

Am Anfang dieser Arbeit steht ein Blick auf vorliegende elektronische Editionen. Dabei sollen Zusammenhänge zwischen der Herstellung elektronischer Editionen und ihren Funktionen hinsichtlich der in ihnen publizierten Texte dargestellt werden. Daran schließt sich die Darstellung prinzipieller Unterschiede beim Umgang mit Texten in Büchern und in elektronischen Medien an. Zuletzt sollen Konzepte zum Umgang mit Texten in elektronischen Medien vorgestellt werden, die den Ansprüchen der germanistischen Arbeit gerecht werden können.

Die Diskussion um elektronische Editionen provoziert häufig unsachliche, ja emotionale Regungen. Aus bibliophilen und anderen Ambitionen heraus, werden elektronische Editionen zum Feindbild erhoben. Der »nüchternen« Bildschirmdarstellung wird entgegengestellt, dass das Lesen in einem »schönen Buch« ein sinnlicher Akt ist, bei dem auch die Materialität des Buches in allen ihren physischen Ausprägungen eine Rolle spielt. Das geht so weit, dass beispielsweise das Geräusch, das beim Umblättern einer Seite entsteht, beim Lesen am Bildschirm vermisst wird – und dass es elektronische Editionen gibt, die beim »Blättern« am Bildschirm solche Geräusche zu simulieren versuchen.

Hier stellt sich die Frage, ob es Sinn und Zweck elektronischer Editionen sein kann, Bücher zu simulieren, oder ob stattdessen nicht andere Aspekte im Vordergrund stehen sollen. Sicherlich wird eine elektronische

Edition nicht das ästhetische Lesevergnügen bereiten, dass ein Buch bereiten kann – von dem Erlebnis, ein altes, wertvolles Buch in Händen zu halten und aufzuschlagen, ganz zu schweigen. Strebt man aber beispielsweise eine Untersuchung zum Wortschatz Goethes an, hält sich das Vergnügen, die 143 Bände der Weimarer Ausgabe nach Belegstellen einzelner Wörter zu durchsuchen, in Grenzen. Um herauszufinden, wie oft Goethe in seinen Briefen das Wort »Geld« benutzt hat, kann man die 50 Bände der vierten Abteilung der Weimarer Ausgabe durchlesen, wofür ein versierter Leser schätzungsweise zwei Wochen benötigt. Eine Volltextsuche im Textbestand der CD-ROM Edition Johann Wolfgang Goethe: Briefe, Tagebücher, Gespräche¹ benötigt etwa zehn Sekunden, um aus dem betreffenden Textbestand 366 Stellen herauszusuchen und aufzulisten. Zusammen mit allen vorkommenden Flexionsformen und Komposita sind es 612 Stellen.

Elektronische Editionen können und sollen Bücher nicht ersetzen, sondern ergänzen. Besondere Vorzüge bieten elektronische Editionen dabei unter folgenden Aspekten:

Wie das oben angeführte Beispiel zeigt, ermöglichen elektronische Editionen einen schnellen Zugriff auf umfangreiche Textbestände. Das Auffinden bestimmter Wörter oder Textstellen ist mit elektronischen Mitteln meist schneller als dies lesend möglich ist. Untersuchungen zum Wortschatz eines Textbestandes lassen sich ebenso schnell realisieren wie beispielsweise statistische Auswertungen von Worthäufigkeiten. Dabei gilt, dass die Möglichkeiten, Texte mit Hilfe elektronischer Medien auszuwerten, um so mächtiger sind, je sorgfältiger eine elektronische Ausgabe besorgt wurde und um so mehr Informationen im elektronischen Text kodiert wurden – was ich im Rahmen dieser Arbeit noch ausführen werde.

Der Inhalt wertvoller Bücher kann über elektronische Medien verfügbar gemacht werden. In vielen Fällen ist es nicht notwendig, auf seltene oder wertvolle Originale zurückzugreifen. Elektronische Editionen können helfen, Rara zu schützen und zu erhalten: Statt ein seltenes Buch den Strapazen der Fernleihe auszusetzen, ist es schonender, es einmal der Beanspruchung der elektronischen Erfassung auszusetzen und zukünftig auf ein elektronisches Faksimile zurückzugreifen.

Da im Fall elektronischer Editionen die Druckkosten entfallen, die bei umfangreichen Werken erheblich sein können, können in elektronischen Medien auch Werke veröffentlicht werden, die aufgrund der Kosten niemals gedruckt werden würden.

Nicht zuletzt geht es auch um die Zukunft der Germanistik. Zum einen ist es notwendig, Möglichkeiten zu entwickeln, die vor dem Hintergrund der Buchkultur erarbeitete Ergebnisse auch über neue Medien zugänglich zu machen – hier stehen vor allem Projekte der elektronischen Erschließung, Retrodigitalisierung und Onlinestellung vorhandener Text- und Wissensbestände im Mittelpunkt. Daneben gilt es, angesichts der neuen Medien und einer neuen Kultur der Kommunikation, den Schritt vom durch Bücher geprägten wissenschaftlichen Diskurs hin zu einer neuen Formen des wissenschaftlichen Austauschs zu wagen. Dies bedeutet besonders für die Geisteswissenschaften, die in viel stärkerem Maße durch die Buchkultur geprägt sind als etwa die Naturwissenschaften, eine große Herausforderung. Da Texte immer mehr im Umfeld elektronischer Medien entstehen, wird sich die Germanistik dieser Medien annehmen müssen, wenn sie sich weiter als Textwissenschaft und nicht als reine Buchwissenschaft verstehen will.

Texte in elektronischen Medien

Die ersten Bemühungen, Computer im Rahmen der germanistischen Arbeit einzusetzen, reichen in die 1960er Jahre zurück. Mikroprozessoren befanden sich noch im Entwicklungsstadium, als elektronische Speichermedien standen Magnetbänder und Lochkarten (s. Abbildung 1) zur Verfügung. In dieser Zeit wurden »Elektronenrechner« bereits auf den Gebieten der Lexikografie und der Wortforschung verwendet, um Wortindizes und Konkordanzen zu umfangreichen Textsammlungen zu erstellen. Dazu mussten die Texte zuerst in eine maschinenlesbare Form

1 Johann Wolfgang Goethe: Briefe, Tagebücher, Gespräche. Berlin: Directmedia Publ., 2000.

gebracht werden, indem sie etwa auf Lochkarten kodiert wurden. Die Listen, die die Rechenanlagen ausgaben, wurden dann wieder von Hand ins Reine getippt, um als Vorlage für den Drucksatz zu dienen.²

Bereits zu dieser Zeit gab es konkrete Vorstellungen über den Einsatz und den Nutzen elektronischer Textsammlungen:

Allmählich verbreitet sich die Ansicht, daß das dynamische Wörterbuch der Zukunft, wie das Spracharchiv selbst [gemeint ist ein Archiv elektronisch gespeicherter Texte – M.M.], als ganzes nie zur endgültigen Fixierung des Druckes gelangen wird, sondern im Speicher der Rechenanlage laufend bereichert sowie durch den einzelnen Benutzer unmittelbar befragt werden wird.³

Auch wenn bereits Texte mit Hilfe der elektronischen Datenverarbeitung (EDV) verarbeitet und Konzepte zum Umgang mit elektronisch gespeicherten Texten entworfen wurden, kam niemand auf den Gedanken, von Texten in elektronischen Medien zu sprechen.



Abbildung 1. Lochkarte mit dem Text der letzten Halbzeile der Nibelungenklage »ditze liet heizt diu klage«⁴

Der Begriff »elektronische Medien« wurde in dieser Zeit meist synonym für die audiovisuellen Medien verwendet. Rundfunk und Fernsehen waren die ersten Medien, die mit Hilfe (analoger) elektronischer Technik Töne und Bilder übermittelten. Texte waren dagegen lange Zeit die Domäne der Printmedien.

Ende der 1970er Jahre entstanden die Voraussetzungen, die einen elektronischen Umgang mit Texten auf der Grundlage der Digitaltechnik auf breiter Ebene ermöglichten. Computer für den persönlichen Gebrauch kamen auf den Markt, es wurden Medien entwickelt, die größere Textmengen speichern (CD-ROM, 1979) oder übertragen (Internet, 1982) konnten. Später folgten Technologien, die darüber hinaus einen komfortablen Umgang mit elektronischen Texten ermöglichten, wie grafische Benutzeroberflächen von Computern (ab Mitte der 1980er Jahre) sowie das World Wide Web und grafikfähige Internetbrowser (1993).

Computermedien und die mit ihnen etablierten Institutionen zur Verbreitung von Daten werden auch als »neue Medien« bezeichnet. Wurden über die neuen Medien zuerst vornehmlich Texte verbreitet, kamen später

2 Vgl. Scholler, Harald: Über die Förderung der Nibelungenforschung durch Elektronenrechner. In: Zeitschrift für deutsches Altertum und deutsche Literatur 25 (1966). S. 149-165; Scholler, Harald: A word index to the Nibelungenklage. Based on the K. Lachmann ed. Ann Arbor: Univ. of Michigan Pr., 1966; Tollenaere, Felician de: Lexikographie mit Hilfe des elektronischen Informationswandlers. In: Zeitschrift für deutsche Sprache 21 (1965), Heft 1. S. 1-19; Wisbey, Roy: Ein Computerlesbares Textarchiv des Frühmittelhochdeutschen. In: Jahrbuch für internationale Germanistik 7 (1969), Heft 2. S. 37-46; Wisbey, Roy A.: Vollständige Verskonkordanz zur Wiener Genesis. Mit einem rückläufigen Wörterbuch zum Formenbestand. Berlin: E. Schmidt, 1967.

3 Wisbey (1969), S. 41.

4 Aus Scholler (1966), S. 152.

Bilder, Töne und Filme bzw. Videos hinzu. Da die »alten« elektronischen Medien zunehmend mit Hilfe digitaler Technik produziert und verbreitet werden und Rundfunk und Fernsehen in absehbarer Zeit in den neuen Medien aufgehen werden, wird der Begriff »elektronische Medien« auch zunehmend für die neuen Medien verwendet und schließt damit auch Texte ein.

Wenn von elektronischen Texten die Rede ist, bedeutet das, dass die Texte in elektronischer Form publiziert werden. Elektronische Daten können von Menschen weder wahrgenommen noch verstanden werden. Spätestens seit Anfang der 1960er Jahre das Magnetband die Lochkarte abgelöst hat, erfolgt die Übertragung und Speicherung elektronischer Daten in Formen, die menschlichen Sinnen unzugänglich sind. Kaum ein Mensch kann den binären Kode, die Grundlage aller digitalen Technologien, unmittelbar verstehen. Viele Ebenen aus Kodes und Protokollen liegen zwischen digitalen Daten, die aus einer Folge der Ziffern »Null« und »Eins« bestehen, und Buchstaben, Bildern oder Tönen, die ein Anwender wahrnehmen und verstehen kann. Computer mit Bildschirm und Lautsprecher, Tastatur und Maus stellen die unverzichtbare Schnittstelle beim Umgang mit elektronischen Daten dar.

Ich erlaube mir im Folgenden, Begriffe wie »Computer« und »elektronische Medien« oder »digitale Medien« weitgehend synonym zu gebrauchen. Letztendlich ist es für einen Anwender, der vor seinem Computer sitzt, kaum noch nachvollziehbar, welche Prozesse auf seinem Arbeitsplatzrechner laufen und welche auf einem Server, mit dem er vernetzt ist, welche Daten, von der Festplatte seines Computers, einer CD-ROM oder aus dem Internet geholt werden. Welche einzelnen elektronischen Speicher- oder Übertragungsmedien zum Einsatz kommen, ist für den Anwender in dem meisten Fällen auch nicht relevant.

1 Elektronische Editionen

Computer werden in unterschiedlichen Bereichen bei der Arbeit an Editionen benutzt. Zum einen werden sie bei der Textarbeit eingesetzt, beispielsweise um Textbestände elektronisch auszuwerten sowie Wortlisten und Konkordanzen zu erstellen; zum anderen werden sie eingesetzt, um Arbeitsergebnisse niederzuschreiben, den Text zu gestalten und den Druck vorzubereiten. Die unterschiedlichen Funktionen, die der Computer bei der Arbeit an Editionen erfüllen kann, spiegeln sich in der Diskussion um elektronische Editionen wider. Bei der grundsätzlichen Frage, inwieweit sich elektronische Medien überhaupt für Editionen eignen, stehen sich zwei unterschiedliche Ansätze gegenüber: der eine ist von den Möglichkeiten und Technologien des Computers geprägt; der andere orientiert sich an Bucheditionen.

Wer von den technischen Möglichkeiten des Computers ausgeht, stellt fest: »Je näher die digitale Version eines solchen Grundlagenwerks [einer Edition; MM] dem Buchmedium steht, umso weniger bietet es mehr als das Buch selbst.«⁵ Dahinter steht der Gedanke, dass die Orientierung am Buch nicht den Möglichkeiten elektronischer Medien gerecht wird. *Textretrieval* und *Hypermedia* sowie der uneingeschränkte Zugriff auf elektronische Ressourcen lassen sich nur konsequent verwirklichen, wenn die Inhalte, um die es geht, auch konsequent vom Informationsträger Papier getrennt und in ein elektronisches System zur Informationsspeicherung⁶ übernommen werden.

Stellt man dagegen das Buch in den Mittelpunkt der Betrachtung, ist gerade die Trennung vom Papier das Manko elektronischer Editionen: »Ungeeignet sind die Bildschirmdarstellungen viel grundlegender deshalb, weil sie den physischen Zusammenhang von Schrift und Schriftträger a priori aufgelöst haben.«⁷ Hier wird der Informationsverlust moniert, der mit der Trennung von Information und ihrer Repräsentation auf dem Medium Papier einhergeht, da »die Materialität der schriftlichen Überlieferung von dieser selbst überhaupt nicht abgelöst werden kann.«⁸ Die elektronischen Möglichkeiten werden demnach nicht denen des Buches gerecht.

1.1 Typen elektronischer Editionen

Betrachtet man vorliegende elektronische Editionen, stellt man fest, dass sich diese beiden Standpunkte in der Praxis widerspiegeln. Es lassen sich grundsätzlich zwei Typen elektronischer Editionen ausmachen.

Der eine Typ elektronischer Editionen löst sich von Darstellungsweisen, wie man sie vom Buch gewöhnt ist. Feste Zeilenlängen oder Seitenumbrüche werden zugunsten einer dynamischen Textdarstellung aufgegeben. Die Gründe hierfür sind zum einen technisch bedingt (Monitorgrößen, Bildschirmauflösungen, skalierbare Programmfenster, verfügbare Schriftarten, Datenhandhabung); zum anderen sind sie elektronischen Editions- und Publikationskonzepten geschuldet. Elektronische Editionen dieses Typs bezeichne ich im Folgenden als »elektronischen Medien nahe stehende elektronische Edition«. Der andere Typ elektronischer Editionen macht im Gegensatz zum Ersten Einschränkungen bei der Nutzung des Potentials elektronischer Medien und hält dafür an den Möglichkeiten und Gepflogenheiten von Bucheditionen fest. Elektronische Editionen dieser Art bezeichne ich im Folgenden als »buchnahe elektronische Editionen«.⁹

5 Burch, Thomas u. a.: Vorwort der Herausgeber. In: Standards und Methoden der Volltextdigitalisierung. Beiträge des internationalen Kolloquiums an der Universität Trier. 8./9. Oktober 2001. Hg. von Thomas Burch u. a. Akademie der Wissenschaften und der Literatur Mainz. Stuttgart: Steiner, 2003. S. 7.

6 Z. B. Filesysteme, Datenbanken und indizierte Textdateien.

7 Reuß, Roland: Textkritische Editionen und Dateiformate. Notizen. In: Jahrbuch für Computerphilologie 1 (1999). S. 101-106. <http://www.textkritik.de/technik/formate.htm> (18.05.2008).

8 Reuß (1999), S. 105.

9 Auf die beiden Editionstypen gehe ich unter 1.3 Buchnahe elektronische Editionen, S. 14 und 1.4 Elektronischen Medien nahe stehende elektronische Editionen, S. 17 noch ein.

Die beiden dargestellten Typen elektronischer Editionen – elektronischen Medien nahe stehende elektronische Editionen und buchnahe elektronische Editionen – haben Modellcharakter. In der Praxis gibt es keine elektronischen Editionen, die vollkommen einem der vorgestellten Konzepte folgen. Die dargestellten Typen sollen lediglich den prinzipiellen Unterschied des editorischen Ansatzes verdeutlichen. In der Praxis ist eine klare Abgrenzung meist schwieriger. Manche Editionen, wie etwa die der DIGITALEN BIBLIOTHEK des Verlags Directmedia Publishing sind vom Konzept und vom Funktionsumfang den elektronischen Medien nahe stehenden elektronischen Editionen zuzuordnen, obwohl sie an einigen von Büchern gewohnten Eigenschaften, etwa einer festen Seiteneinteilung bei der Textdarstellung, festhalten. Andere Editionen dieses Typs, bei denen der ursprüngliche Textzusammenhang des Buchs zugunsten einer höheren elektronischen Funktionalität aufgelöst wurde, enthalten den Text zusätzlich in einem Format, das einen festen Seitenaufbau beibehält. Die Edition der MITTELHOCHDEUTSCHEN WÖRTERBÜCHER¹⁰ bietet beispielsweise die zugrunde gelegten Wörterbücher nicht nur in einer mediengerechten Form, sondern auch als Dateien im *Portable Document Format (PDF)*, das häufig bei buchnahen elektronischen Editionen verwendet wird.

Auf der Seite der buchnahen elektronischen Editionen muss angemerkt werden, dass sich *PDF*-Dateien bei entsprechender Auslegung elektronisch durchsuchen lassen, und damit einfache Funktionen des *Datenretrievals*¹¹ ermöglichen. Daneben gibt es Editionen, bei denen die bildliche Textdarstellung einer Image-Edition¹² mit Möglichkeiten des elektronischen Textzugriffs verbunden wurde, wie etwa im Rahmen der sehr aufwendigen HISTORISCH KRITISCHEN GOTTFRIED KELLER AUSGABE¹³ oder der elektronischen Ausgabe des UNIVERSAL-LEXICONS von Zedler, das im Rahmen eines DFG-Projekts von der Bayerischen Staatsbibliothek digitalisiert wurde.¹⁴ Auch die Google-Buchsuche, die es ermöglicht, Inhalte digitalisierter Bücher im Internet zu durchsuchen und einzusehen, lässt sich diesem Typ zuordnen.¹⁵

Auffällig ist, dass viele buchnahe elektronische Editionen als Hybrideditionen im Zusammenhang mit Bucheditionen erscheinen. Als Hybridausgaben werden allgemein Editionen bezeichnet, die aus einem Buch und einem elektronischen Medium, meist einer CD-ROM, bestehen. Dabei enthält der elektronische Datenträger entweder den Textbestand des Buches oder ergänzendes Material zum Buch.

Neben den funktionalen Unterschieden zwischen elektronischen Medien nahe stehenden elektronischen Editionen und buchnahen elektronischen Editionen lassen sich Unterschiede beim Editionsstand der Texte ausmachen. Während bei den elektronischen Medien nahe stehenden elektronischen Editionen weniger komplexe Texteditionen vorliegen – meist mit dem Niveau einer »kommentierten Leseausgabe« vergleichbar –, bieten buchnahe elektronische Editionen oft anspruchsvollere Textausgaben.

Ich werde im Folgenden genauer auf buchnahe und elektronischen Medien nahe elektronische Editionen eingehen. Dabei werde ich zunächst die Hintergründe der beiden Typen darstellen. Zum einen möchte ich damit eine Erklärung für den oben konstatierten Status quo elektronischer Editionen liefern, zum anderen werde ich aufzeigen, welche Aspekte im Hinblick auf andere Konzepte für elektronische Editionen, wie ich sie im weiteren Verlauf der Arbeit vorstellen werde, berücksichtigt werden müssen. Aus dem Dargestellten werde ich Ansätze

10 Burch, Thomas u. a. (Hg.): *Mittelhochdeutsche Wörterbücher im Verbund*. CD-ROM und Begleitbuch. Stuttgart: Hirzel, 2002.

11 Vgl. 1.4.4 Retrieval, S. 23.

12 Vgl. Image-Editionen, S. 20.

13 Keller, Gottfried: *Sämtliche Werke*. Hg. von Walter Morgenthaler u. a. Basel u. a.: Stroemfeld u. a., 1996 ff.

14 Zedler, Johann Heinrich: *Grosses vollständiges Universal-Lexicon aller Wissenschaften und Künste*. Leipzig 1732–1754. <http://mdz10.bib-bvb.de/~zedler/zedler2007/index.html> (18.05.2008).

15 <http://www.google.de/books> (18.05.2008).

und Vorschläge entwickeln, wie und auf welcher Ebene anspruchsvolle elektronische Editionen verwirklicht werden können. Dabei werde ich mich an folgenden Fragestellungen orientieren:

Welche Informationen enthalten Texte auf Papier?

In diesem Zusammenhang betrachte ich Text als »aufgeschriebene Sprache« und gehe davon aus, dass (gesprochene) Sprache ursprünglich linear (eindimensional) ist. Diesen Ansatz möchte ich zum einen benutzen, um die verschiedenen Informationsebenen darzustellen, die ein Text neben der reinen Sprachinformation enthält; zum anderen ergibt sich hier der kleinste gemeinsame Nenner zu elektronischen Medien, bei denen Informationen sequentiell übertragen werden.

Durch das Aufschreiben von Sprache wird die Sprache »materialisiert«. Die lineare Sprache wird dabei um verschiedene Ebenen ergänzt, die in ihrer Gesamtheit das schriftsprachliche Gewebe eines Textes ausmachen.

- Aufgeschriebene Texte können durch Leerräume wie Abstände, spezielle Trennzeichen, Absätze und Seitenumbrüche strukturiert und durch Überschriften gegliedert werden.
- Schrift ermöglicht die Nutzung verschiedener Schriftarten, Schriftschnitte, Schriftgrößen und Schriftauszeichnungen. Durch Typografie und Layout kommen dem Text neue Informationsebenen zu.
- Zeichen und Wörter oder Absätze können über-, unter- oder nebeneinander geschrieben werden. Papier ist ein zweidimensionales Medium.
- Der Befund des verwendeten Schreibmaterials (Papier, Stift/Maschine) birgt Informationen über die Schreibsituation des Autors und den Kontext des Textes.
- Gesprochene Sprache ist ein transitorisches Medium, Papier ein kontinuierliches. Papier kann als »Zeitachse« betrachtet werden. Korrekturen und Veränderungen hinterlassen Spuren.
- Durch seine Materialität in Form von Papier und Buch hat ein Text einen materiellen Kontext, der Auskunft über verschiedene Fragen erteilen kann, zum Beispiel: Wo und wie wurde er verbreitet? Wo steht er (in welcher Zeitschrift, welchem Sammelband, welcher Bibliothek)?

Wie geht die Editionswissenschaft mit diesen Ebenen von Text um?

In vielen Fällen begnügt sich die Editionswissenschaft damit, die Befunde der Textarbeit lediglich zu transkribieren. Die Handschrift eines Autors wird in gedruckte Buchstaben überführt und Besonderheiten der Vorlage typografisch markiert, etwa, indem in der Handschrift unterstrichene Passagen im Druck kursiv wiedergegeben werden. Die Editionswissenschaft bleibt dem Medium Papier verhaftet. Auch wenn die einzelnen Befunde im Arbeitsprozess abstrahiert werden und in einer Fußnote, einem Anhang oder sonstigen Apparat Niederschlag finden, ist das Endziel doch die Erstellung einer neuen Form des Textes. Einige Informationen, etwa Quellenbeschreibungen, werden separiert und in einen eigenständigen Text umgesetzt; andere Befunde werden lediglich »umformatiert«. Informationen, die dem Text entnommen werden können, ohne bereits in den Bereich der Interpretation zu fallen, werden nicht expliziert. Hat der Autor eines Dramas im Manuskript die Sprecher unterstrichen, werden sie in einer Druckausgabe vielleicht kursiv wiedergegeben. Die Bedeutung der Textauszeichnung ist den meisten Lesern klar. Wer den Text aufgrund seiner Kenntnisse der typografischen Merkmale¹⁶ als Drama erkennt, weiß, dass es sich bei hervorgehobenen Wörtern um die Sprecher der folgenden Textzeilen handelt. Der Befund wird also umkodiert – von einer Unterstreichung zu einer Kursivierung –, aber nicht erklärt. Im Kontext schriftlicher Texte erscheint eine explizite Angabe, dass es sich bei kursivem Text um einen Sprecher (oder eine Regieanweisung etc.) handelt. Wie relevant solche Informationen im Kontext elektronischer Editionen sind, werde ich noch aufzeigen.

16 Vgl. Entwicklung typografischer Gliederungsmerkmale, S. 44.

Welche Möglichkeiten bieten elektronische Medien?

Betrachtet man die Arbeitsweise elektronischer Medien, stellt man zunächst fest, dass es Analogien zu gesprochenen Texten gibt.¹⁷ Der Datenstrom, der die Grundlage für komplexe elektronische Darstellungsmöglichkeiten bildet, ist ursprünglich linear – analog zur linear gesprochenen Sprache. Entsprechend finden sich einige Techniken, mit denen geschriebene Sprache gegliedert wird, bei der Gliederung und Strukturierung von Daten wieder. So wurde das phonetische Alphabet im Laufe der Schriftentwicklung um besondere Zeichen ergänzt, die keinen Lautwert haben, beispielsweise Satzzeichen, Leerzeichen oder Anführungszeichen.

Auch zur Strukturierung von Daten werden spezielle Zeichen verwendet. Bei *Auszeichnungssprachen*¹⁸ wie *Hyper Text Markup Language (HTML)*, *Standard Generalized Markup Language (SGML)* oder *Extensible Markup Language* finden sich viele Methoden, die analog zu denen der Gliederung von Texten betrachtet werden können. So werden Leerzeichen zur grundsätzlichen Gliederung in »Sinneinheiten« verwendet; spezielle Ausnahmezeichen bekommen eine konkrete Bedeutung.

Die Gemeinsamkeit zwischen geschriebenem bzw. gedrucktem Text und elektronischen Medien endet da, wo medienspezifische Eigenschaften eine Rolle spielen. Während man auf Papier etwa eine bestimmte Schriftauszeichnung zur Kennzeichnung einer Textstelle einsetzt und somit eine spezifische Möglichkeit der Textformatierung verwendet, muss im Falle elektronischer Medien eine Art »Zwischenstufe« eingefügt werden. Die endgültige, »feste« Ausgabe einer Information, indem sie etwa ausgedruckt wird, bildet im Bereich der elektronischen Medien lediglich eine von mehreren Möglichkeiten, Informationen auszugeben. Die Stärke elektronischer Medien liegt gerade darin, Informationen flexibel handhaben zu können, um etwa verschiedene Darstellungsweisen zu ermöglichen. Im Gegensatz zum »festen« Buchsatz einer Edition benötigen elektronische Medien Texte in einer Art »flüssigem« Aggregatzustand. Erst die Nutzung bestimmter Technologien generiert dann eine mögliche feste Repräsentationsform der jeweiligen Daten.

Wie können elektronische Medien für literaturwissenschaftliche Editionen verwendet werden?

Hier schließt sich der Bogen zur Ausgangsproblematik. Aus den vorangegangenen Überlegungen möchte ich Vorschläge entwickeln, wie Texte von Seiten der Editionswissenschaft aufgearbeitet und mit welchen Technologien sie mit Hilfe elektronischer Medien für den wissenschaftlichen Gebrauch nutzbar gemacht werden können. Dabei sind Ansätze von Seiten der geisteswissenschaftlichen Disziplin, etwa die *Text Encoding Initiative* oder die *Dublin Core Metadata Initiative (DC bzw. DCMI)*, ebenso zu berücksichtigen wie Technologien zum Umgang mit elektronischen Texten wie sie besonders im Umfeld von *XML*- und *X-Media Publishing*-Konzepten entwickelt wurden.

Am Ende kann dabei kein Universalrezept zur Erstellung elektronischer Editionen stehen –, so wie es kein solches Konzept für Editionen in Buchform gibt und geben kann. Ziel dieser Arbeit ist es, Überlegungen vorzustellen, mit denen man die anfangs dargestellte Kluft zwischen den beiden Positionen in Bezug auf elektronische Editionen verringern kann.

1.2 Entstehungsbedingungen elektronischer Editionen

Ausgangspunkt dieser Arbeit ist die Beobachtung, dass sich die bisher erschienenen elektronischen Editionen grob in zwei Kategorien einteilen lassen: dem Buch nahe stehende elektronische Editionen und elektronischen Medien nahe stehende elektronische Editionen. Im Folgenden werde ich detaillierter auf die beiden Typen elektronischer Editionen eingehen und ihre Entstehungsbedingungen, die an sie gerichteten Anforderungen und die

¹⁷ Vgl. 2.2.1 Speicherung von Texten in elektronischen Medien, S. 59.

¹⁸ Vgl. 3.1 Auszeichnungssprachen, S. 71.

mit ihnen realisierten Funktionen darstellen. Ich werde aufzeigen, dass die beiden unterschiedlichen Konzepte aus den unterschiedlichen Arbeitsvoraussetzungen heraus zu erklären sind, die ihrer Entstehung zugrunde liegen.

Während dem Buch nahe stehende Editionen meist im Rahmen neuer Editionsprojekte herausgegeben werden, entstehen elektronischen Medien nahe stehende elektronische Editionen häufig auf der Grundlage vorhandener, abgeschlossener Ausgaben. Arbeitet man bei der Editionserstellung mit Computern, ist das Arbeitsziel in den meisten Fällen die Herausgabe eines Buches, so dass die Editionsarbeit darauf ausgerichtet ist. Das Ziel der editorischen Arbeit ist ein Buch, eine »elektronische Version« des Textes lediglich ein Nebenprodukt. Der Computer wird dabei vornehmlich als Hilfsmittel bei der Fertigung des papiernen Endproduktes betrachtet. Dies heißt, dass man sich bei der Arbeit vornehmlich auf der Ebene des virtuellen Papiers mit dem Computer beschäftigt.

Bei elektronischen Medien nahe stehenden Editionen ist meist eine vorliegende Buchedition Ausgangspunkt der Arbeit. Das Arbeitsziel ist, den Inhalt der Buchausgabe möglichst vielfältig für den Zugriff mit Hilfe elektronischer Medien nutzbar zu machen. Hier tritt der Computer als Medium in den Mittelpunkt des editorischen Interesses. Abbildung 2 veranschaulicht die unterschiedlichen Arbeitsprozesse.

Die unterschiedlichen Entstehungsbedingungen der beiden Konzepte elektronischer Editionen scheinen zunächst unlogisch –, sollte man doch meinen, dass aktuelle Editionsprojekte, bei denen große Teile der Editionsarbeit EDV-unterstützt geleistet werden, zu elektronischen Medien nahe stehenden Editionen führen. Schließlich werden bei der Editionsarbeit elektronische Medien eingesetzt: Texte und Bilder liegen bei der Editionserstellung bereits in elektronischen Datenformaten vor. Genauso unverständlich scheint es, dass viele den elektronischen Medien nahe stehende Editionen auf der Grundlage älterer, abgeschlossener Ausgaben entstehen. Hier müssen Texte und Materialien erst durch Techniken zur Retrodigitalisierung¹⁹ in ein elektronisches Format gebracht werden.

Ziel dieser Darstellung ist es, die eingangs dargestellte Beobachtung, dass es für elektronische Editionen noch kein tragfähiges editorisches Konzept gibt, anhand des aktuellen Umgangs mit elektronischen Medien im Bereich der Edition zu untermauern und gleichzeitig Ansatzpunkte aufzuzeigen, die ein entsprechendes Konzept berücksichtigen muss.

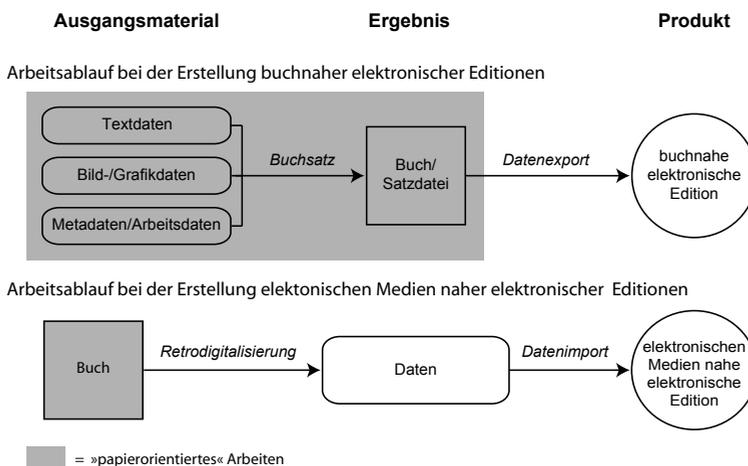


Abbildung 2. Schematische Darstellung des Arbeitsablaufes bei der Erstellung elektronischer Editionen

19 Vgl. I.4.7 Retrodigitalisierung, S. 29.

1.3 Buchnahe elektronische Editionen

Elektronische Editionen, die gedruckten Editionen nahe stehen, werden meist erst am Ende des Arbeitsprozesses an einer neuen Edition erstellt, wenn die eigentliche Editionsarbeit geleistet wurde und die Daten für den Druck erzeugt werden. Wie zu früheren, vor-elektronischen Zeiten, als die entsprechend vorbereiteten Texte dem Setzer übergeben wurden, gehört dieser Arbeitsschritt – die Erzeugung der Satzdaten – oft nicht mehr in den unmittelbaren Verantwortungsbereich des Herausgebers. Fällt die Erstellung der elektronischen Edition in diese Phase des Arbeitsablaufs, obliegt sie demnach auch nicht mehr der Betreuung durch den Herausgeber, der ggf. den Text für seine Verwendung in elektronischen Medien einrichten könnte. Die für eine Buchedition konzipierten Texte werden lediglich in ein weiteres Datenformat umgesetzt, das nicht primär zum Druck, also zur Ansteuerung von Belichtungsmaschinen bestimmt ist, sondern zur Darstellung des Textes in elektronischen Medien. Viele aktuelle Satzprogramme bieten Funktionen, um neben Satzdateien auch andere Datenformate zu erzeugen. Diese können dann mit geeigneten Programmen am Computer genutzt werden. Als geeignete Programme haben sich zum einen Webbrowser, zum anderen der Adobe Reader durchgesetzt.

Webbrowser, wie beispielsweise der Microsoft Internet-Explorer oder Mozilla Firefox, wurden ursprünglich dazu entwickelt, Textseiten, die in *Hyper Text Markup Language (HTML)* kodiert wurden, darzustellen und *Hyperlinks* sowohl innerhalb eines Textes als auch zwischen verschiedenen Dokumenten im Internet zu ermöglichen. Durch die rasante Weiterentwicklung von Technologien für das *World Wide Web (WWW, W3)* bieten heutige Browser weitere Funktionen. Da Browserprogramme speziell zur Darstellung von Texten und Bildern entwickelt wurden, stellen sie bereits viele Möglichkeiten zur Verfügung, um auf entsprechend eingerichtete elektronische Editionen zurückgreifen können. Neben reinen *HTML*-kodierten Texten, die aufgrund von Einschränkungen bei der Darstellung formatierter Texte für elektronische Editionen kaum verwendet werden, kommt bei aktuellen elektronischen Editionen *Extensible Markup Language* zum Einsatz, wobei die Darstellung der *XML*-kodierten Texte durch den Einsatz von *Stylesheets*²⁰ weitgehend vorgegeben werden kann. Viele Satzprogramme können Satzdaten in einer geeigneten Form – *HTML*- oder *XML*-kodiert – ausgeben.

Adobe Reader und Adobe Acrobat²¹ sind die zurzeit meistbenutzten Programme, wenn es um die Bereitstellung und Darstellung von Texten in elektronischen Medien geht. Der frei für die meisten Betriebssysteme verfügbare Adobe Reader dient der Anzeige entsprechender Dokumente, Adobe Acrobat zu ihrer Erzeugung.²² Das zugrundeliegende Datenformat ist das 1993 von Adobe entwickelte *Portable Document Format*. Es dient dem plattform- und systemübergreifenden Austausch von Dokumenten. In erster Linie gewährleistet das *PDF* die originalgetreue Wiedergabe eines Dokuments auf verschiedenen Computersystemen und in unterschiedlichen Ausgabemedien. Dokumente im *PDF* können dabei für einzelne Verwendungen optimiert werden. Die Möglichkeiten reichen von der Optimierung für die Verbreitung über das Internet und die Darstellung am Bildschirm bis zur Druckvorstufe, wo es andere Datenformate, beispielsweise *PostScript*, weitgehend abgelöst hat.

Als Beispiel einer solchen Edition möchte ich hier die Ausgabe *FRANZ KAFKA: HISTORISCH-KRITISCHE AUSGABE SÄMTLICHER HANDSCHRIFTEN, DRUCKE, TYPOSKRIPTE*²³ anführen. Die den Buchbänden beiliegenden CD-ROMs enthalten den Textbestand und die faksimilierten Kafka-Handschriften der Buchausgabe im *PDF*. Zum Zugriff auf die Texte benötigt der Nutzer den *Adobe Reader*. Die Darstellung der elektronischen Version entspricht den

²⁰ Vgl. Technologien zur Formatierung, S. 92.

²¹ Die Namen der Programme hat Adobe öfter geändert. Frühere Versionen hießen Acrobat Professional und Acrobat Reader.

²² Mittlerweile gibt es auch von anderen Firmen Programme zur Erzeugung von *PDF*-Dokumenten.

²³ Kafka, Franz: *Historisch-kritische Ausgabe sämtlicher Handschriften, Drucke, Typoskripte*. Hg. von Roland Reuf. Basel/Frankfurt a. M.: Stroemfeld, 1995 ff.

Seiten der Buchausgabe. Dabei folgt die elektronische Version keinem eigenen Konzept. Lediglich die einfache Suchfunktion des *Adobe Readers* erweitert den Nutzwert über das reine Lesen am Bildschirm hinaus.

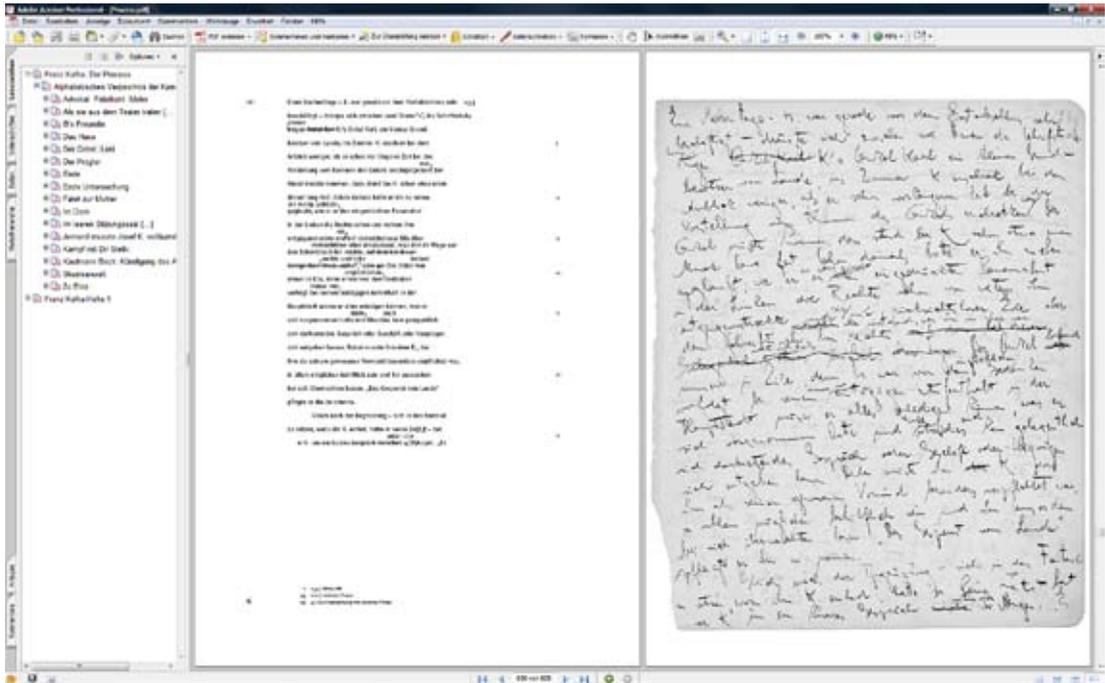


Abbildung 3. Screenshot der PDF Version von Franz Kafka: Historisch-kritische Ausgabe sämtlicher Handschriften, Drucke, Typoskripte. Hg. von Roland Reuß. Basel/ Frankfurt a. M.: Stroemfeld, 1997

Die meisten buchnahen elektronischen Editionen entstehen im Rahmen neuerer Editionsprojekte, bei denen elektronische Medien bei der Editionsarbeit zum Einsatz kommen, hier spricht man gern von »elektronischem Edieren«.

Um zu verdeutlichen, wie diese Form elektronischer Editionen zustande kommt und wie bei der Arbeit an Papiereditionen Computer eingesetzt werden, skizziere ich im Folgenden die grundsätzlichen Schritte der EDV-unterstützten Editionsarbeit. Die erste Phase bei der Arbeit an einer Edition besteht in den meisten Fällen aus der Sammlung und kritischen Sichtung des vorliegenden Materials. Auf dem Gebiet der EDV spricht man in diesem Fall von der Datenerhebung bzw. Datenerfassung. Die zweite Phase stellt die Auswertung des Materials dar. Obwohl es bei verschiedenen Editionsprojekten erhebliche Unterschiede bei Gegenstand, Methode und Editionsziel gibt, möchte ich diese Phase pauschal als die der Datenverarbeitung bezeichnen. Das Ziel der Editionsarbeit ist schließlich die Publikation der Edition, die Datenausgabe.²⁴

24 Vgl. z. B. Ott, Wilhelm: Computer-unterstützte Editionen. In: editio. Internationales Jahrbuch für Editions-wissenschaft. Bd. 3 (1989). S. 157-176. Ott geht sehr detailliert auf die bei der Edition erforderlichen Arbeitsschritte sowie auf geeignete Programme ein. Angesichts der heute zur Verfügung stehenden Programme und der damit veränderten Möglichkeiten für die Editionsarbeit möchte ich bewusst ein gröberes Schema entwerfen.

1.3.1 Datenerfassung

Bei der Arbeit an Editionen ist es notwendig, das zu edierende Material sowie benötigte Hintergrundinformationen zusammenzutragen. Da diese Arbeit weitgehend mit Hilfe des Computers geleistet wird, bezeichne ich diesen Schritt als Datenerfassung. Die Datenerfassung kann verschiedene Ebenen der Materialsammlung und -sichtung umfassen. Zunächst bietet es sich an, die vorhandenen Textzeugen mit einer Beschreibung und weiteren relevanten Daten in einer Datenbank zu sammeln.²⁵ Nach Sichtung und Bewertung der Zeugen – die hier gewonnenen Ergebnisse können ebenfalls elektronisch gespeichert und ausgewertet werden – wäre der nächste mögliche Schritt, das eigentliche Material, z. B. den Text oder einzelne Zeugen, in eine elektronische Form zu überführen. Je nach Material und Editionsziel bieten sich dazu verschiedene Verfahrensweisen an. Texte werden meist mit dem Computer erfasst bzw. in ein Datenformat übertragen, in dem eine Weiterverarbeitung mit Computern möglich ist. Dazu kommen häufig Textverarbeitungsprogramme, Datenbankprogramme oder auch Bildbearbeitungsprogramme zum Einsatz.

Liegen die einer Edition zugrunde liegenden Zeugen als seltene Originale, etwa in Form von Manuskripten, Typoskripten oder auch Tonaufzeichnungen vor, wird die elektronische Erfassung meist von Experten vorgenommen. Hier kann die Kenntnis einer Handschrift, das Wissen um spezifische Arbeitsweisen eines Autors oder die Erfahrung in der Transkription von Tonaufzeichnungen für die Datenerhebung unverzichtbar sein. Auch Besonderheiten – etwa Korrekturen oder sonstige bedeutsame Befunde – können so gleich bei der Texterfassung berücksichtigt werden.

In Fällen, in denen eine verlässliche Grundlage existiert, die keine spezifischen Kenntnisse bei der Erfassung für die EDV voraussetzt, greift man auf weniger aufwendige Verfahren der Datenerfassung zurück, wie sie etwa bei der Retrodigitalisierung Anwendung finden.²⁶

1.3.2 Datenverarbeitung

Datenerfassung, Datenverarbeitung und Datenausgabe lassen sich nicht isoliert betrachten. Bereits bei der Datenerfassung muss berücksichtigt werden, wie die weitere Verarbeitung erfolgen soll und welche Daten dabei benötigt werden. Hier kommen wieder unterschiedliche Programme zum Einsatz: Textverarbeitungs-, Datenbank- und Bildverarbeitungsprogramme genauso wie Satzprogramme.

Spezialisierte Programme wie etwa TUSTEP (Tübinger System von Textverarbeitungsprogrammen), dessen Grundlagen bereits 1966 am Deutschen Rechenzentrum Darmstadt entwickelt wurden,²⁷ oder spezialisierte Datenbank-Programme unterstützen die Arbeit an wissenschaftlichen Editionen auf breiter Ebene.

1.3.3 Datenausgabe

Die erhobenen Daten werden letztendlich an ein Satzprogramm übergeben. Die praktischen Arbeitsabläufe variieren hierbei stark. Teilweise können die erarbeiteten Texte gleich in komplexer Form in das Satzprogramm übernommen werden – indem etwa Anmerkungen, Apparate oder Register, die mit anderen Programmen erstellt wurden, in geeigneter Form in das Satzbild einfließen. Teilweise werden sie erst im Satzprogramm in ihrer endgültigen Form erstellt.

²⁵ Die Praxis zeigt, dass hierfür bevorzugt ein Tabellenkalkulationsprogramm, beispielsweise Excel, eingesetzt wird.

²⁶ Vgl. 1.4.7 Retrodigitalisierung, S. 29.

²⁷ Auf der Homepage von Tustep <http://www.zdv.uni-tuebingen.de/tustep/index.html> (18.05.2008) finden sich umfangreiche Bibliografien mit Arbeiten zum Edieren mit Tustep sowie ein Verzeichnis mit Tustep realisierter Projekte. Weitere Informationen finden sich auf der Homepage der International Tustep User Group (ITUG) unter <http://www.itug.de/> (18.05.2008). Vgl. auch Castrillo Benito, Nicolás und Stahl, Peter (Hg.): TUSTEP educa. Actas de Congreso del International TUSTEP User Group. Penaranda de Duero (Burgos) Octubre 1999. Burgos: Univ., 2001. Sowie Reeg, Gottfried und Schubert, Martin J. (Hg.): Edieren in der elektronischen Ära. Berlin: Weidler, 2004.

In den meisten Fällen wird der Drucksatz nicht mehr auf der Grundlage eines mit Markierungen versehenen Typoskripts erstellt,²⁸ sondern auf der Grundlage einer formatierten Datei, die mit Hilfe eines Textverarbeitungsprogramms erstellt wurde. Die Ebene des »papiernen Denkens« wird hier nicht verlassen.

Neben den Satzdaten für den Druck, werden aus dem Satzprogramm auch häufig Daten für andere Ausgabemedien generiert – etwa im *PDF* oder als *HTML/XML*-kodierte Daten für das Internet oder für eine CD-ROM-Edition. Da auch hier meist die Darstellung im Vordergrund der Arbeit steht, werden selten tiefer gehende elektronische Möglichkeiten berücksichtigt.

1.4 Elektronischen Medien nahe stehende elektronische Editionen

Bei elektronischen Medien nahe stehenden elektronischen Editionen ist in den meisten Fällen das Buch nicht Arbeitsziel sondern Ausgangspunkt einer elektronischen Edition. Die Berücksichtigung medienpezifischer Möglichkeiten und Anforderungen sind hier essentieller Teil des Editionsconzeptes. Solche Editionen werden oft auf der Grundlage bereits vorhandener Bucheditionen erstellt. Bei der dazu meist erforderlichen Retrodigitalisierung kann man bereits Aspekte, die für eine elektronische Ausgabe relevant sind, berücksichtigen. Aspekte, die bei Editionen für den Druck bedacht werden müssen, können dagegen entfallen, da die Buchausgabe bereits vorliegt.

Als Beispiel für eine elektronischen Medien nahe stehende Edition möchte ich die elektronische Kafka-Edition *KAFKAS WERKE IM WWW* bzw. *KAFKAS WERKE AUF CD-ROM*²⁹ anführen.³⁰ Die im Internet und auf CD-ROM-Datenbank verfügbare elektronische Edition enthält den Text der Buchausgabe *FRANZ KAFKA, KRITISCHE AUSGABE, SCHRIFTEN UND TAGEBÜCHER*³¹ samt Apparaten.

Die Funktionen der CD-ROM Datenbank sind etwas umfangreicher als die der im Internet verfügbaren Version. Die Anzeige des Textes auf dem Bildschirm erlaubt eine flexible Darstellungsbreite; die Seitenumbrüche der Vorlage sind erkennbar. Verweise auf den Apparat sind als *Hyperlinks* realisiert. Der Zugriff auf die Texte erfolgt über ein hierarchisches Inhaltsverzeichnis, das der Struktur der Bände der Buchausgabe folgt. An Suchmöglichkeiten steht eine Volltextsuche zur Verfügung, die gängige Suchmöglichkeiten erlaubt: Einschränkung auf verschiedene Bereiche, Boolesche Operatoren, Trunkierungen.³² Zur Nutzung der komplexeren Suchfunktionen muss der Nutzer eine spezielle Syntax erlernen. Das Ergebnis einer Suchabfrage wird als Fundstellenliste ausgegeben. Die einzelnen Fundstellen können nacheinander aufgesucht werden, sie werden im Text markiert. Suchabfragen und Fundstellenlisten lassen sich abspeichern. Text kann markiert und exportiert werden.

Abbildung 4 zeigt die im Internet verfügbare Edition *KAFKAS WERKE IM WWW*. Die Grundlage bildet ein mit *SGML* ausgezeichnete Text.³³ Die Textdarstellung verzichtet auf komplexe Formatierungen. Dies gewährleistet, dass der Text auf vielen verschiedenen Computersystemen richtig dargestellt werden kann. Ergänzende Informationen sind mit Hilfe von *Hyperlinks*,³⁴ im Text der Abbildung rot, verknüpft. So können komplexe

28 Vgl. Beispiel 1. Textauszeichnung in einem Typoskript, S. 72.

29 Kafka, Franz: *Kafkas Werke im WWW/Kafkas Werke auf CD-ROM*. Kritische Kafka-Ausgabe des S. Fischer Verlages bei Chadwyck-Healey. Chadwyck-Healey, 1999.

30 Die elektronische Ausgabe ist das Produkt einer Zusammenarbeit des S. Fischer Verlags mit Chadwyck-Healey.

31 Kafka, Franz: *Kritische Ausgabe. Schriften, Tagebücher, Briefe*. Hg. von Jürgen Born u. a. Frankfurt a. M.: Fischer, 1982 ff.

32 Vgl. 1.4.4 Retrieval. S. 23.

33 Vgl. Beispiel 2. Auszug aus dem SGML-Quellcode der elektronischen Editionen *KAFKAS WERKE IM WWW* und *KAFKAS WERKE AUF CD-ROM*. S. 75 f.

34 Vgl. 1.4.3 Hypertext und Hypermedia, S. 22.

Informationen zum Text textnah verfügbar gemacht werden, ohne auf umfassende typografische Hilfsmittel zurückgreifen zu müssen.

Habe ich oben dargestellt, inwieweit »buchnahe« elektronische Editionen dadurch geprägt werden, dass sie dem »papiernen Denken« verhaftet bleiben, da bei der editorischen Arbeit mittels Computer als Arbeitsziel eine Buchausgabe im Vordergrund steht, werde ich nun aufzeigen, wie elektronischen Medien nahe stehende Editionen im Wesentlichen durch Funktionen und Möglichkeiten von elektronischen Medien geprägt werden. Orientieren sich Konzepte und Überlegungen buchnaher Editionen an den Möglichkeiten des Buchdrucks, orientieren sich elektronischen Medien nahe stehende Editionen im Unterschied dazu an elektronischen Medien. Entsprechend werde ich die Möglichkeiten zum Umgang mit Text, die elektronische Medien bieten, darstellen. Dazu werde ich die wesentlichen Funktionen elektronischen Medien nahe stehender Editionen erläutern und darstellen, wie durch Retrodigitalisierung elektronische Editionen auf der Grundlage von Bucheditionen entstehen. Daneben soll das Dargestellte Grundlagen vermitteln, die notwendig sind, um zu verstehen, welche Anforderungen an ein Editions-konzept zu stellen sind, das elektronische Medien als Publikationsmedium von vornherein berücksichtigt.



Abbildung 4. Screenshot Kafkas Werke im WWW. Kritische Kafka-Ausgabe des S. Fischer Verlages bei Chadwyck-Healey. Chadwyck-Healey, 1999

1.4.1 Navigation

Der Begriff Navigation wird im Bereich der EDV verwendet, wenn es darum geht, Informationen durch die Nutzung vorgegebener Strukturen aufzusuchen. Beispiele für solche Strukturen sind Inhaltsverzeichnisse, Register und Indizes oder Wortlisten. Das Gegenmodell zur Navigation ist die Informationssuche mit Hilfe des so genannten Information *Retrieval*, bei dem ein Textbestand nach vom Benutzer definierten Kriterien durchsucht und die Ergebnisse auf verschiedene Arten ausgegeben werden können.³⁵

35 Vgl. 1.4.4 Retrieval, S.23.

Im Rahmen elektronischer Editionen kommt der Navigation insofern eine besondere Bedeutung zu, als dass sie eine von Büchern her gewohnte Herangehensweise bietet. Eine entsprechend ausgelegte Navigationsfunktion ermöglicht es dem Nutzer, sich auch in großen Textbeständen zurechtzufinden. Dies ist besonders wichtig, da elektronische Editionen sehr umfangreich sein können. Die elektronische Goethe-Ausgabe von Chadwyck-Healey³⁶ etwa versammelt den vollständigen Text der 143 Bände der Weimarer Ausgabe in einer Datenbank; die CD-ROM-Edition Deutsche Literatur von Lessing bis Kafka³⁷ enthält auf ca. 160.000 Seiten die Werke von mehr als 100 Autoren (das entspricht in etwa 12 Regalmetern Bücher.) Ein Nutzer dieser Ausgaben steht aber nicht vor einem Buchregal, das ihm die einzelnen Bände räumlich strukturiert zugänglich macht, sondern er sitzt vor einem Bildschirm, auf dem er die gewünschten Texte »ansteuern« muss.

Die einfachste Form der Navigation ist das lineare »Durchblättern« eines Textbestandes. Allerdings unterstützen nicht alle elektronischen Editionen ein solches Herangehen. Die Möglichkeit des Blätterns setzt eine Art Seiteneinteilung des elektronischen Textes voraus. Dies ist aber gerade bei elektronischen Medien nahe stehenden elektronischen Editionen eher die Ausnahme. Meist wird der Text in einer Form dargestellt, die an den Bildschirm bzw. die Möglichkeiten der Bildschirmdarstellung angepasst ist. Bei Texten, die über die Darstellungsmöglichkeiten des Bildschirms hinausgehen, muss der Benutzer daher »scrollen«, d. h. sich mit Hilfe der entsprechenden Tasten der Computertastatur oder der Maus im Text auf und ab bewegen. Bei umfangreichen Texten verliert man hier schnell die Übersicht. Umso wichtiger ist es, dass elektronische Medien geeignete Navigationshilfen anbieten, damit sich ein Leser im Textbestand orientieren kann.

Wie bei gedruckten Editionen bilden Inhaltsverzeichnisse und Register bzw. Indizes die häufigsten Formen, um in umfangreichen Textbeständen den gezielten Zugriff auf einzelne Texte, Textstellen oder Informationen zu ermöglichen. Inhaltsverzeichnisse werden vielfach in Form von Strukturbäumen dargestellt, wie sie auch bei Computern zur Darstellung der Verzeichnisstruktur von Ordnern und Dateien benutzt werden. Durch diese wird gleichzeitig die Hierarchie der einzelnen Kapitel und Absätze deutlich. Diese Form der Darstellung ist insofern von weitergehender Bedeutung, als viele Technologien³⁸ zur Adressierung von Daten auf einer solchen, »Baumstruktur« genannten Struktur aufbauen.³⁹ Ein Vorteil der Baumstruktur ist, dass der Nutzer in vielen Fällen bestimmen kann, bis zu welcher Gliederungstiefe sie angezeigt wird. Die meisten elektronischen Editionen bieten diese Form der Navigation an, wobei die Darstellung natürlich variieren kann. Die einzelnen Einträge eines Strukturbaumes bezeichnet man technisch als »Knoten«. Sie bilden Knotenpunkte für die Navigation.

Problematisch wird diese Darstellung allerdings, wenn eine Ebene ausfällt, d. h. auf einen Absatz der dritten Gliederungsebene unmittelbar einer der fünften folgt. Der Knoten der vierten Gliederungsebene würde hier fehlen, was im Zusammenhang mit vielen gängigen Technologien zu Problemen führt.

Register und Indizes fallen bei verschiedenen elektronischen Editionen dagegen sehr unterschiedlich aus – angefangen von einfachen Listen aller im elektronischen Text vorkommenden Wörter bis hin zu hierarchischen und thesaurierenden Indizes. Elektronische Indizes verfügen in den meisten Fällen über eine *Hypertextfunktion*, d. h. ein Klick auf ein Wort im Index führt zur Anzeige der betreffenden Textstelle. Indizes, die nicht redaktionell aufbereitet, sondern automatisch aus dem Text generiert wurden, bieten in vielen Fällen die Möglichkeit, den Kontext, in dem das jeweilige Wort steht, anzuzeigen.

36 <http://goethe.chadwyck.com/> (18.05.2008).

37 Bertram, Mathias (Hg.): Deutsche Literatur von Lessing bis Kafka. Studienbibliothek. Berlin: Directmedia Publishing, 2000.

38 Im Kontext von XML sind etwa XPath und XPointer zu nennen.

39 Vgl. etwa 3.2.3 Darstellungsweisen, S. 86 und Technologien zur Adressierung, S. 93.



Abbildung 5. Ausschnitt aus dem Inhaltsbaum der CD-ROM-Edition *Goethes Werk im Kontext*. Die »Berliner Ausgabe« auf CD-ROM. Berlin: Karsten Worm, 2005

1.4.2 Textdarstellung

Die essentielle Funktion aller elektronischen Editionen ist die Anzeige des Textbestandes. Bei dieser Art der Textdarstellung muss grundsätzlich unterschieden werden, ob es sich bei dem dargestellten Text um Text- oder Bilddaten handelt. Die Unterscheidung ist insofern problematisch, als letztendlich jedes am Bildschirm dargestellte Zeichen grafisch, also als Bild, angezeigt wird und ein Benutzer nicht auf den ersten Blick unterscheiden kann, welche Art von Daten dem angezeigten Text zugrunde liegen. Da dieser Unterschied zum Verständnis der Funktionen elektronischer Editionen wichtig ist, werde ich ihn im Folgenden kurz erläutern.

Volltext-Editionen

Liegen der elektronischen Edition Textdaten zugrunde, spricht man von einer Volltext-Edition. Die Inhalte einer solchen Edition liegen in Form von Zeichendaten vor. Ein »a« ist in einer Form gespeichert, durch die der Computer weiß, dass es sich um den Buchstaben »a« handelt. Ein solches »a« kann über die Tastatur eingegeben werden, entsprechende Programme können in einem Text nach einem solchen »a« suchen, ein solches »a« kann aus einem Text heraus kopiert und in einen anderen eingefügt werden. Das Kopieren und Einfügen von Text ist auch zwischen unterschiedlichen Programmen möglich, so lange sie mit Zeichendaten arbeiten. Das Aussehen des »a«, also die Form, in der es am Bildschirm angezeigt wird, wird von der Schriftart bestimmt, die dem Buchstaben zugewiesen ist. Dazu sind den meisten elektronischen Texten Informationen mitgegeben, die für den Leser unsichtbar sind und die festlegen, in welcher Schriftart der Buchstabe oder der ganze Text auf dem Bildschirm dargestellt wird. Die Information, wie der Buchstabe konkret aussehen soll, d. h., welches Bild der Computer für ein »a« in einer bestimmten Schriftart auf dem Bildschirm anzeigt, wird durch die Schriftart bestimmt. Schriftarten müssen auf dem jeweiligen Computer vorhanden sein und liegen hier meist als so genannte *Fonts* vor. Neuere Technologien, die über das Internet auf zentral gespeicherte *Fonts* zurückgreifen, konnten bisher aus rechtlichen und technischen Gründen nicht realisiert werden.

Image-Editionen

Im Gegensatz zu Volltext-Editionen enthalten so genannte Image-Editionen nur die Informationen über das Aussehen des Textes. Ein »a« wird vom Computer nicht als einzelner Buchstabe »a« gespeichert, sondern als Teil eines Bildes, das den ganzen Text als ein Bild darstellt. Für den Benutzer ist eine Image-Edition von einer Volltext-Edition an verschiedenen Punkten zu unterscheiden. Vergrößert man etwa die Darstellung bei einer Volltext-Edition, werden die Buchstaben auch in der Vergrößerung klar und scharf dargestellt. Der Computer berechnet

die Darstellung jedes einzelnen Buchstaben optimal für jede Darstellungsgröße. Vergrößert man dagegen ein Image, wird das Gesamtbild vergrößert und die Pixel des Bildes werden sichtbar.



Abbildung 6. Bild eines »a« in vergrößerter Darstellung



Abbildung 7. Buchstabe »a« in großer Darstellung

Bei einer Image-Edition ist nicht jeder einzelne Buchstabe ein kleines Bild, sondern meist wird eine ganze Textseite als ein Bild angelegt. Entsprechend sind bei der Darstellung keine Veränderungen beim Zeilen- oder Seitenumbruch möglich. In den meisten Fällen lässt sich der Text nicht markieren oder in andere Anwendungen übernehmen. Allerdings gibt es Techniken, die solche Funktionen auch im Fall von Image-Editionen erlauben.

Verwendung von Volltext- und Image-Editionen

Volltext-Editionen finden sich meist unter den elektronischen Medien nahe stehenden Editionen. Zum einen ist eine flexible Textdarstellung beim Umgang mit elektronischen Medien vorteilhaft, weil es sehr unterschiedliche Hardwarevoraussetzungen, wie etwa verschiedene Bildschirmgrößen, gibt. Zum anderen, weil elektronische Medien es dem Benutzer vielfach erlauben, die Darstellung den eigenen Bedürfnissen bzw. der eigenen Arbeitsumgebung anzupassen. Daneben ist die Möglichkeit, Text in andere Programme zu übernehmen und weiter zu verwenden ein genereller Vorteil des Arbeitens in elektronischen Medien. Auch lassen sich viele Funktionen elektronischer Editionen leichter mit einer Volltext-Edition realisieren.

Gegen Volltext-Editionen spricht, dass vorliegende Textbestände unter Umständen erst retrodigitalisiert werden müssen. Wie oben erwähnt, liegen elektronischen Medien nahen elektronischen Editionen häufig etablierte Buchausgaben zugrunde, bei denen selten auf vorhandene Daten zurückgegriffen werden kann. Eine Retrodigitalisierung stellt aber einen nicht unerheblichen Arbeitsaufwand dar.⁴⁰ Daneben müssen für Volltext-Editionen eigene editorische Konzepte entwickelt werden, da man hier meist nicht auf die von Büchern her gewohnten Darstellungsweisen zurückgreifen kann. Diese beruhen zum großen Teil auf einem festgelegten Seitenlayout und einer exakt definierten Typografie.⁴¹

Image-Editionen halten auf Kosten der optimalen Anpassungsmöglichkeit an elektronische Medien an einem starren Seitenaufbau und einer genau vorgegebenen Typografie fest. Sie bieten das Abbild einer Buchseite. Je nach Anspruch der elektronischen Edition kann die Retrodigitalisierung des Textes und somit ein aufwendiger Arbeitsgang entfallen oder mit einem geringeren Aufwand betrieben werden, da ein Benutzer immer das Bild des Originals vor Augen hat. Typografische Eigenheiten spielen in diesem Fall ebenso eine untergeordnete Rolle wie gelegentliche Fehler bei der Texterfassung. Ein weiterer Vorteil ist, dass das Editions-konzept einer Buchausgabe nahezu unverändert übernommen werden kann, da Typografie und Seitendarstellung nicht veränderlich sind. Feste Text- und Seitenumbrüche gewährleisten etwa die gewohnte Möglichkeit des Textnachweises durch die Angabe von Seitenzahlen. Daneben kommen sie den Gewohnheiten aus dem Umgang mit Büchern entgegen: Die Orientierung im Text fällt bei einer festen Seitendarstellung meist leichter, da man sich an der Seitenstruktur orientieren kann. Zur Darstellung des edierten Textes stehen die vom Buchsatz her gewohnten Möglichkeiten zur Verfügung, besonders wenn auch komplexere Textzusammenhänge wie Textgenese oder Textvarianten dargestellt werden sollen.

⁴⁰ Vgl. 1.4.7 Retrodigitalisierung, S. 29.

⁴¹ Vgl. 2.1 Texte in Büchern, S.38.

Anmerkungen und Apparate

Buchnahe elektronische Editionen richten sich bei der Darstellung von Anmerkungen und Apparaten meist nach den von Büchern gewohnten Darstellungsweisen. Anmerkungen oder Apparate stehen unten auf der jeweiligen »Seite« oder am Ende des Textes. Um dem Nutzer das »Blättern« in der elektronischen Ausgabe zu ersparen, werden Text und Anmerkung/Kommentar meist mittels *Hyperlinks* verknüpft.

Bei elektronischen Medien nahe stehenden Editionen sind die Möglichkeiten zur Darstellung von Apparaten vielfältiger. So kann der betreffende Apparat etwa automatisch parallel zum Text in einem eigenen Bildschirmbereich angezeigt, per »mouse over«⁴² temporär eingeblendet oder natürlich ebenfalls per Hyperlink aufgerufen werden.

1.4.3 Hypertext und Hypermedia

Die ursprüngliche Idee von *Hypertext* ist es, der normalen, linearen Informationsstruktur innerhalb herkömmlicher Texte einen assoziativen Zugang zu Informationen gegenüberzustellen. Dieser ermöglicht zum einen die weitgehend redundanzfreie Darstellung komplexer Zusammenhänge. Zum anderen geht man davon aus, dass assoziative Informationszusammenhänge dem menschlichen Denken näher sind als lineare. Die Entwicklung von *Hypertext*-Konzepten wird meist auf den von Vannevar Bush im Jahr 1945 veröffentlichten Artikel »As We May Think«⁴³ und das 1960 von Ted Nelson initiierte Projekt »Xanadu«⁴⁴ zurückgeführt.

Bush legt dar, dass wissenschaftliche Unterlagen (records) in Bibliotheken über ein linear geordnetes, hierarchisch gegliedertes Indexsystem zugänglich sind, während der menschliche Verstand dagegen assoziativ arbeitet. (Wissens-)Einheiten (items) werden durch den menschlichen Verstand in ein Netz von Pfaden (web of trails) eingebunden. Bush entwirft das Konzept einer Maschine, die er »Memex« (memory extender) nennt und die als Erweiterung des Gedächtnisses dienen soll. Eine Schlüsselfunktion der Maschine sollte, neben der Eingabe, Abfrage und parallelen Darstellung von Informationen die Möglichkeit sein, einzelne Einheiten über Pfade zu verknüpfen sowie eigene Kommentare und Anmerkungen einzufügen. Die Pfade sollen es zum einen ermöglichen, beim Aufruf einer Information auch die bereits verbundenen Informationen aufzurufen; zum anderen sollen Pfade und Anmerkungen zwischen verschiedenen Memex-Maschinen austauschbar und damit von anderen Wissenschaftlern benutzbar sein.

Auch wenn der Entwurf seiner Memex im Wesentlichen noch auf der Technik von Mikrofilmen beruht, hat Bush bereits die wichtigsten Grundlagen moderner Datenorganisation formuliert: die Organisation der Informationen in einem Netz, in dem sie über Pfade verknüpft werden können.

Nelson wird die Prägung des Begriffs *Hypertext* zugeschrieben.⁴⁵ Ziel seines seit 1960 verfolgten und bis heute weiterentwickelten Projekts »Xanadu« war die Initiierung eines Universums von Dokumenten (docuverse) in elektronischer Form. Xanadu ist ein komplexes System, das neben einem nicht linearen Textzugang als wesentliche Bestandteile auch die Entstehungs- und Überarbeitungsgeschichte eines jeden Textes protokolliert sowie

42 »Mouse over« heißt, dass der Computer etwas ausführt, wenn der Mauszeiger nur über einer bestimmten Stelle stehen bleibt, ohne dass der Nutzer eine weitere Aktion, etwa einen Mausklick, ausführt.

43 Bush, Vannevar: As We May Think. In: The Atlantic Monthly, 7/1945 (Vol. 176, Heft 1). S. 101-108.

44 Vgl. <http://xanadu.com> (18.05.2008).

45 Nelson beschäftigte sich seit 1960 mit der Entwicklung von Software zur Verwaltung von Texten. Den Begriff Hypertext hat er angeblich seit 1963 verwendet; der erste schriftliche Nachweis findet sich 1965 in Wedeles, Lauren: Prof. Nelson Talk. Analyzes P.R.I.D.E. In: Vassar Miscellany News, 3.2.1965, S. 3 f. Hierbei handelt es sich um einen Bericht über einen Vortrag Nelsons mit dem Titel »Computers, Creativity and the Nature of the Written Word«. Der Begriff bekam durch sein Buch LITERARY MACHINES große Popularität. Vgl. Nelson, Theodor Holm: Literary machines. The report on and of project Xanadu concerning word processing, electronic publishing, hypertext, thinkertoys, tomorrow's intellectual revolution and certain other topics including know-ledge, education and freedom. Sausalito, CA: Mindful Press, 1982.

ein Abrechnungssystem umfasst. Auch wenn sich »Xanadu« nicht auf breiter Ebene durchsetzen konnte, beruhen viele aktuelle Technologien im Bereich des Informationsretrievals auf Konzepten Nelsons.

Unser heutiges Verständnis des Begriffs *Hypertext* ist durch seine Verwendung im Kontext des *World Wide Web* geprägt. Die für die Verwendung im *WWW* entwickelte Auszeichnungssprache *Hyper Text Markup Language* trägt den Begriff *Hypertext* bereits im Namen.

Die Grundidee von *Hypertext* wird hier auf die Verwendung so genannter »*Hyperlinks*« beschränkt. *HTML* erlaubt es, von einer Textstelle auf eine andere Textstelle oder ein anderes Dokument zu verweisen. Im Gegensatz zu komplexeren *Hypertext*-Konzepten gelten dabei folgende Einschränkungen:

- *Hyperlinks* müssen bereits durch den Autor eines *Hypertextes* angelegt werden und sind vom Leser nicht zu beeinflussen.
- *Hyperlinks* führen nur in eine Richtung; die Rückkehr zum Ausgangstext muss entweder ebenfalls durch den Autor des *Hypertextes* angelegt oder über die Softwarefunktion des Browsers (History-Funktion) realisiert werden.

So gesehen sind *Hyperlinks* nichts anderes als die elektronische Umsetzung von Noten, Querverweisen oder Marginalien in Büchern, wie es sie bereits seit dem Mittelalter gibt. Lediglich das Suchen der betreffenden Textstelle wird durch einen Mausklick ersetzt.

Hypermedia bezeichnet in diesem Kontext ebenfalls nichts wirklich Neues: Wird der Begriff *Hypermedia* verwendet, will man meist zum Ausdruck bringen, dass sich Verweise nicht nur auf Texte beziehen können, sondern auch auf andere, multimediale Daten wie Bilder oder Audio- und Videodokumente. Auch solche Verweise wurden im Zusammenhang mit Büchern bereits realisiert, nur sind die Möglichkeiten beim Einsatz des Computers bequemer nutzbar, da alle Datenformen auf einem Gerät wiedergegeben werden können. Das Konzept ist nicht neu: Illustrierte Bücher, in denen auf Abbildungen verwiesen wird; Bücher, zu denen es ergänzendes Material auf einem Tonträger gibt (etwa auf Audiokassette, Schallplatte oder Audio-CD).

Die Struktur von *Hypertexten* scheint dem menschlichen Denken jedoch nicht so nah zu sein, wie es die Theorie vorsieht. So ist in der bisherigen *Hypertext*-Praxis auffällig, dass zur Informationssuche doch immer wieder auf lineare Strukturen zurückgegriffen wird, etwa auf so genannte »guided tours«, bei denen der »Leser« eines *Hypertextes* der vom Autor vorgeschlagenen Lesestruktur folgen kann oder durch eine grafische Aufarbeitung der Informationsanordnung in Form eines Plans oder Baumes.

Inwieweit dem Aspekt der Assoziativität Rechnung getragen wird, wenn in der Praxis alle relevanten Verknüpfungen bereits als »Links« vom Autor des *Hypertextes* angelegt sein müssen und somit nur der Assoziativität des Autors, in anderen Fällen der Redaktion oder bei modernen Content Management Systemen (CMS) der Software entsprechen, sei dahingestellt. Hier gibt es noch umfangreiches Entwicklungspotential, das sich im Spannungsfeld zwischen variabel einsetzbaren Linksammlungen, intelligenten Suchmaschinen und Konzepten eines Semantic Web bewegt.

1.4.4 Retrieval

Der Begriff *Retrieval* oder ausführlich »Information Retrieval« stammt aus den Bereichen der Informationswissenschaft und der Informatik. Er wird meist mit dem deutschen Wort »Informationswiedergewinnung« übersetzt. Im Rahmen elektronischer Editionen bezeichnet man mit *Retrieval* die freie Suche nach Buchstaben, Wörtern und Wortfolgen im Textbestand. Das *Retrieval* stellt damit eine der mächtigsten Funktionen elektronischer Editionen dar. Durch die Formulierung geeigneter Suchanfragen kann ein Nutzer gezielt Textstellen auffinden oder sich bedarfsgerechte Register in Form von Listen, etwa Wortlisten oder

Fundstellenlisten, zusammenstellen lassen. Was bei gedruckten Editionen nur durch Lesen des gesamten Textbestandes möglich ist – sofern für die Edition nicht eine umfangreiche Konkordanz erstellt wurde –, ermöglichen moderne *Retrieval*-Funktionen in Sekunden. Das *Retrieval* geht damit einen Schritt über die Möglichkeiten hinaus, die die Navigation bietet, und ermöglicht den Zugriff auf den Text, ohne dass ein Benutzer auf vorgegebene Strukturen zurückgreifen muss, wie es bei der Navigation der Fall ist.

Um *Retrieval*-Funktionen nutzen zu können, müssen zwei Voraussetzungen erfüllt sein. Erstens muss der Nutzer wissen, welche Informationen er sucht, und zweitens muss er in der Lage sein, eine geeignete Suchanfrage zu formulieren.

Die erste Voraussetzung, dass der Benutzer wissen muss, welche Informationen er sucht, scheint zunächst evident. Doch gerade hier gibt es einige Probleme, die gerade einem Nutzer, der den lesenden Umgang mit Texten gewohnt ist, nicht unbedingt augenfällig sind. Während ein Leser etwa in den meisten Fällen problemlos abweichende Wortformen, Schreibweisen oder Textfehler erkennt, hält sich eine einfache elektronische Suche strikt an eine vorgegebene Buchstabenfolge. Weiß ein Nutzer nicht, welche Rechtschreibung dem Text zugrunde liegt, ob er nach »Zierath« (Adelung), »Zierat« (Grimm, »alte« deutsche Rechtschreibung) oder »Zierrat« (neue deutsche Rechtschreibung) oder gar nach einer Schreibweise in einer ganz anderen Sprachstufe oder einem Dialekt suchen muss, wird er unter Umständen nicht fündig werden. Um solche Probleme zumindest teilweise abfangen zu können, sollte ein Nutzer in der Lage sein, eine geeignete Suchanfrage zu formulieren.

Bei der Unterstützung, eine geeignete Suchanfrage zu formulieren, zeigen sich gravierende Unterschiede zwischen verschiedenen elektronischen Editionen. Das Spektrum reicht von komfortablen Suchmasken, die dem Benutzer bei der Formulierung einer geeigneten Suchanfrage helfen, bis hin zu einer simplen Kommandozeile, auf der die Suchabfrage mit Suchbefehlen in einer entsprechenden Syntax eingegeben werden muss.

Im Folgenden werde ich auf einige *Retrieval*-Funktionen eingehen, die bei den meisten elektronischen Medien nahe stehenden elektronischen Editionen zur Verfügung stehen. Leider lässt sich in Bezug auf bestehende elektronische Editionen feststellen, dass die heute technisch möglichen Such- bzw. *Retrieval*-Funktionen bei weitem nicht ausgeschöpft werden. Mit der Zunahme elektronisch vorliegender Texte werden auch andere Technologien zur Informationsauswertung an Bedeutung gewinnen, beispielsweise so genannte »Text Mining«-Technologien.⁴⁶

Trunkierungen

Viele Suchfunktionen realisieren eine Suche unter Verwendung so genannter Trunkierungen, Platzhalterzeichen oder »Wildcards«. Dabei steht ein bestimmtes Zeichen für ein oder mehrere andere Zeichen. Bekannt ist etwa der Stern »*«, der für eine beliebige Menge beliebiger Zeichen steht. Eine Suche nach »Editio*« würde sowohl »Edition« als auch »Editionen« aber auch »Editionswissenschaft« finden. Das Fragezeichen »?« steht meist als Platzhalter für genau einen beliebigen Buchstaben. »?aum« findet demnach etwa »Baum«, »Saum« aber nicht »Traum«. Daneben gibt es, je nach Edition, noch verschiedene andere Trunkierungsmöglichkeiten, etwa »#« für eine beliebige Ziffer. Manche Suchfunktionen erlauben auch anzugeben, welche Zeichen im Bereich des Platzhalters vorkommen dürfen: »[d-g]« kann etwa bedeuten, dass an dieser Stelle die Buchstaben »d«, »e«, »f« und »g« stehen können.

Trunkierungen eignen sich gut, um unterschiedliche Wortformen zu finden. Allerdings lassen sich entsprechende Suchanfragen schwer präzise formulieren. Hilfreich ist dabei, wenn die Suchfunktion es ermöglicht, Wortlisten mit allen gefundenen Wörtern zu erstellen. Eine Suchanfrage über einen umfangreichen Textbestand⁴⁷

46 Vgl. Mehler, Alexander: Textmining. In: Texttechnologie. Perspektiven und Anwendungen. Hg. von Lothar Lemnitzer und Henning Lobin. Stauffenburg, Tübingen, 2004, S. 329-352.

47 Als Beispiel habe ich die CD-ROM Deutsche Literatur von Lessing bis Kafka – Studienausgabe herangezogen. Diese elektronische Edition enthält Werke von 108 Autoren seit der Mitte des 18. Jahrhunderts bis in die 20er Jahre des 20. Jahrhunderts in teils vollständiger, teils

nach allen vorkommenden Formen des Verbs »gelten«, die leichtfertig mit »g?lt*« formuliert wurde, ergab über 800 verschiedene Wörter.

Boolesche Operatoren

Eine weitere Möglichkeit stellt die Suche mit so genannten Booleschen Operatoren dar: Mehrere Suchausdrücke können durch die Operatoren UND, ODER, NICHT und XOR (ausschließendes oder) verknüpft werden. Mit dem UND-Operator können etwa mehrere Wörter in einem Zusammenhang gesucht werden; ODER ermöglicht es, alternative Wortformen zu suchen. NICHT erlaubt es, bestimmte Zusammenhänge auszuschließen. XOR sucht dagegen den einen oder den anderen Ausdruck, nicht aber das gleichzeitige Vorkommen beider Ausdrücke. In vielen Zusammenhängen lassen sich Boolesche Operatoren nur sinnvoll einsetzen, wenn sich die Reichweite der Suchanfrage beschränken lässt, etwa auf eine bestimmte Anzahl von Wörtern, auf Sätze, Absätze oder Kapitel.

Unschärfe Suche

Unter den Begriffen »unscharfe Suche«, »Fuzzy-Suche« oder »tolerante Suche« versteht man im Allgemeinen, dass bei der Suche vorgegebene Toleranzkriterien berücksichtigt werden. So kann z. B. auch eine Suche nach phonetisch gleichwertigen Wörtern ermöglicht werden oder Texte aus verschiedenen Sprachstufen einfach durchsucht werden.

Um eine unscharfe oder tolerante Suche zu ermöglichen, gibt es verschiedene Konzepte. Eine Möglichkeit besteht darin, eine definierte Abweichung zuzulassen. So kann man etwa erlauben, dass ein Buchstabe im zu findenden Wort von dem gesuchten Wort abweicht. Die Angabe, wie groß die Abweichung vom Suchbegriff sein darf, kann auf verschiedene Weisen geschehen. So kann etwa festgelegt werden, wie viele Buchstaben vom Suchbegriff abweichen dürfen oder wie viel Prozent Abweichung erlaubt sind. Entsprechende Toleranzkriterien sollten im Idealfall auf den zu durchsuchenden Text abgestimmt sein, damit typische Schreibweisenvarianten sicher erfasst werden. Bei dieser Form der unscharfen Suche werden oft nur unbefriedigende Ergebnisse erzielt. Sucht man etwa das Wort »sein« mit einer Toleranz, die einen abweichenden Buchstaben zulässt, würde man auch die Variante »seyn« finden, aber etwa auch »seit« etc. Um einen solchen Fall aufzufangen, würde sich die Suche mit Trunkierungen (»se?n«) oder mit Booleschen Operatoren (»sein ODER seyn«) anbieten.

Eine andere, häufig genutzte Möglichkeit ist es, bei der Suche Listen zu berücksichtigen, die Angaben mit erlaubten Abweichungsgraden enthalten. Diese können etwa häufige Schreibweisenvarianten erfassen, beispielsweise »i« = »y«, »f« = »ph« etc. oder phonetisch gleichwertige Buchstabengruppen. Gibt ein Nutzer ein Wort ein, das ein »i« enthält, sucht die Software auch alle Wörter, die an der gleichen Stelle ein »y« enthalten. Mit der Suche nach »sein« wird dann auch »seyn« gefunden. Noch aufwändiger ist es, der Suche ein Lexikon zu hinterlegen, das abweichende Wortformen, etwa bei der Flexion von Verben, verzeichnet. Sucht ein Nutzer beispielsweise nach »gehen«, sucht eine entsprechende Software zunächst im Index nach »gehen« und findet hier alle Wortformen, die dann im Text gesucht werden, beispielsweise »geht«, aber auch »ging«. Der Aufbau eines solchen Lexikons ist aufwändig. In den meisten Fällen werden komplette Wörterbücher zum Aufbau solcher Lexika einbezogen. Der Vorteil für den Benutzer ist, dass bei einer Suchanfrage quasi zuerst im Wörterbuch nachgesehen wird, welche grammatischen Formen es von einem Wort gibt und diese dann alle gesucht werden. Allerdings funktioniert diese Technik nur bei Texten mit einheitlicher Orthografie zuverlässig.⁴⁸ Im Falle von älteren Textbeständen mit uneinheitlicher Orthografie ist es möglich, spezifische Wörterbücher zu hinterlegen. Probleme können etwa bei mehrsprachigen Texten auftreten.

repräsentativer Auswahl, zusammen ca. 170 000 Seiten.

48 Ein recht neues Problem stellt z. B. die Rechtschreibreform von 1996 dar, deren Umsetzung 2006 abgeschlossen sein sollte.

Hinterlegte Informationen

Es gibt Fälle, in denen eine Information, die man in einem Text finden möchte, nicht explizit, in einer stark abgewandelten Form oder in so unterschiedlichen Formen vorkommt, dass sie mit gängigen Suchmethoden nicht gefunden werden kann. Einige Beispiele:

Problematisch für die Erschließung eines Textes mit gängigen *Retrieval*-Funktionen kann etwa sein, wenn der Text Eigenheiten der Orthografie eines Autors, Mundarten oder Schreibfehler übernimmt. Wenn dem Nutzer diese speziellen Eigenschaften des Textes oder auch nur von Textpassagen nicht bekannt sind, können unter Umständen relevante Textstellen nicht gefunden werden.

Ähnliche Probleme können bei fremdsprachigen Passagen auftreten, besonders wenn der Text zusätzlich in einem anderen Alphabet geschrieben ist. Text, der in einem anderen Schriftsystem geschrieben wurde, kann in elektronischen Medien meist noch dargestellt werden. Problematisch ist es aber, wenn man auf einer normalen Tastatur einen griechischen, hebräischen oder gar japanischen Text eingibt, nach dem man suchen möchte. In diesem Fall ist es hilfreich, wenn dem dargestellten Text Informationen hinterlegt sind, etwa, wie er in lateinische Buchstaben transkribiert bzw. transliteriert zu finden ist. Die Transkriptions-/Transliterationsregeln müssen dem Nutzer dazu natürlich bekannt sein. Auch unterschiedliche Transkriptionen können ein Problem verursachen, etwa wenn in einem Textbestand unterschiedliche Transkriptionsregeln verwendet wurden. Entweder muss der Herausgeber hier redaktionell eingreifen oder die Informationen müssen versteckt hinterlegt werden. Ein »normaler« Nutzer kommt vielleicht nicht auf den Gedanken, dass er, wenn er bei der Suche nach »Tschechov« fündig geworden ist, besser auch noch nach »Tschechow« oder gar »Čechov« sucht (von »Чехов« einmal ganz abgesehen).

Auch bei Bildsymbolen und Bildern kann hinterlegter Text hilfreich sein. In seinen Tagebüchern verwendet Victor Klemperer an einigen Stellen im Text ein Hakenkreuz, wenn er das Wort »nationalsozialistisch« meint. Die Bedeutung wird beim lesenden Zugang zum Text sofort klar, für die elektronische Suche sollte ein entsprechendes Wort hinterlegt werden, da eine Suchanfrage die betreffende Stelle sonst nicht finden kann.

Im Fall von Image-Editionen (s. o.) liegen ganze Texte als Bilder vor. Hier ist es sinnvoll, den abgebildeten Seiten Text zu hinterlegen. Je nachdem, um was für einen Text es sich handelt, genügen einzelne Schlüsselwörter. In der ersten Phase der elektronischen Edition des ZEDLER⁴⁹ wurden beispielsweise nur die Stichwörter hinterlegt. Über die Suche nach Stichwörtern waren die passenden Artikel auffindbar. Die Artikel selber lagen lediglich als Image vor und konnten am Bildschirm gelesen werden.⁵⁰

Sucheinschränkungen

Sucheinschränkungen können das *Retrieval* auf verschiedenen Ebenen unterstützen. Einige elektronische Editionen bieten die Möglichkeit, die Suche auf bestimmte Texte oder Textbereiche einzuschränken. Dies ist sinnvoll, wenn der Textbestand einer elektronischen Edition sehr umfangreich ist, die Suche aber nur in einem bestimmten Teilbestand erfolgen soll.

Für den sinnvollen Einsatz einer Suche mit Booleschen Operatoren ist es wichtig, den Suchbereich vom Umfang her auf einzelne Abschnitte, Kapitel, Sätze oder auf noch kleinere Strukturen einzuschränken. Eine Suche, die die Wörter »Knab« und »Röslein« mit einem Booleschen UND verbindet, wird nur sicher zum Heideröslein führen, wenn die Suche auf einen Bereich von wenigen Wörtern eingeschränkt wird. Ansonsten gibt es sicher einige Texte, in denen die beiden Wörter vorkommen.

Eine andere Möglichkeit von Sucheinschränkungen besteht darin, in Unterstrukturen zu suchen. Bei Literaturanthologien kann es sinnvoll sein, die Suche auf eine bestimmte Literaturform einzuschränken zu können,

49 Zedler, Johann Heinrich: Grosses vollständiges Universal-Lexicon aller Wissenschaften und Künste. Leipzig, Halle: 1731–1754.

50 Vgl. <http://mdz10.bib-bvb.de/~zedler/zedler2007/index.html> (18.05.2008).

etwa auf Gedichte. In Dramen oder bei Transkripten mündlicher Texte⁵¹ kann die Möglichkeit nützlich sein, nur im Sprechtext einer bestimmten Person zu suchen.

Statistische Suchmöglichkeiten

Neben den aufgeführten Möglichkeiten des *Retrievals* bieten einige elektronische Editionen eine statistische Suche. Diese Suchmöglichkeit wird auch als Wortfeldsuche oder Themensuche bezeichnet. Der Benutzer kann eine Liste von Begriffen erstellen, die für das Thema, zu dem im Text recherchiert werden soll, relevant sind. Der Text wird dann nach allen in der Liste enthaltenen Wörtern durchsucht. Als Ergebnis einer solchen Suche wird ausgegeben, wie hoch die Vorkommenshäufigkeit der jeweiligen Wörter in einzelnen Textbereichen ist. Der Nutzer bekommt damit einen Anhaltspunkt, welche Textstelle am ehesten die von ihm gewünschten Informationen enthält. Bei dieser Form der Suche werden in den elektronischen Editionen unterschiedliche statistische Algorithmen eingesetzt. Dies macht die Ergebnisse solcher Suchformen zwischen verschiedenen elektronischen Editionen schwer vergleichbar.

Ergebnisanzeige

Neben den Suchfunktionen an sich kommt der Ergebnisanzeige im Rahmen des *Retrievals* eine wichtige Bedeutung zu. Eine übersichtliche Ergebnisanzeige sollte es dem Nutzer ermöglichen, auf einen Blick zu erkennen, inwieweit die bzw. welche Suchergebnisse für ihn relevant sind oder ob er die Suchanfrage optimieren kann. Auch hier lassen sich zwischen den einzelnen elektronischen Editionen große Unterschiede ausmachen.

Im einfachsten Fall wird der Text Stück für Stück durchsucht. Jedes Auslösen des Suchvorgangs bewirkt, dass die jeweils nächste Fundstelle im Text angezeigt wird. Das gefundene Wort wird dabei meist durch so genanntes »Highlighting« – die meist farbige Hinterlegung der gefundenen Stelle – markiert. Diese Form der Suche ist etwa mit der Suchfunktion des Textverarbeitungsprogramms Microsoft Word vergleichbar. Wie das Gesamtergebnis der Suchanfrage aussieht, erfährt man nur, wenn man sich alle Fundstellen einzeln ansieht.

Wesentlich komfortabler sind Programme, die zuerst die Suche ausführen und das Ergebnis zunächst in Form einer Fundstellenliste anzeigen. So hat ein Nutzer zumindest einen Überblick darüber, wie viele relevante Stellen gefunden wurden. Aus der Fundstellenliste kann man meist gleich zur betreffenden Textstelle springen.

Die Fundstellenliste ist als Arbeitsmittel umso wertvoller, je mehr Informationen sie über die Fundstelle enthält. Stehen die meisten oben beschriebenen Suchfunktionen bei nahezu allen gängigen elektronischen Editionen mehr oder weniger komfortabel zur Verfügung, gibt es bei der Ergebnisanzeige die größten Unterschiede. Um nicht alle Möglichkeiten einzeln abzuhandeln, stelle ich an dieser Stelle die Fundstellenliste der DIGITALEN BIBLIOTHEK des Verlags Directmedia Publishing vor. Die DIGITALE BIBLIOTHEK bietet eine der im Vergleich umfassendsten Fundstellenlisten. Dabei werden die Ergebnisse nicht nur als Liste, sondern in Form einer Tabelle dargestellt, die sich wiederum nach den ergänzenden Angaben sortieren oder filtern lässt.

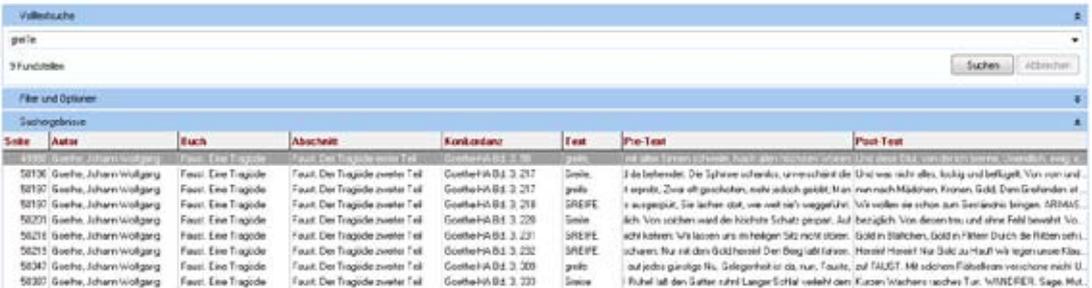
Für die Abbildung 8 wurde im Band 1, Deutsche Literatur von Lessing bis Kafka innerhalb der FAUST Dramen Goethes nach dem Suchausdruck »grei?e« gesucht. Die abgebildete Ergebnisliste enthält folgende Angaben: In der ersten Spalte erfolgt die Angabe, auf welcher »Seite«⁵² der digitalen Ausgabe der gefundene Text steht. Klickt man mit der Maus eine Zeile der Liste an, wird im gegenüberliegenden Bildschirmbereich gleich die entsprechende Seite angezeigt und die betreffende Fundstelle hervorgehoben. Die nächsten drei Spalten geben an, in welchem Bereich des Textbestandes die Fundstelle liegt. Im Fall des als Beispiel herangezogenen Bandes, der nach Autoren gegliedert ist, sind dies »Autor«, »Buch« und »Abschnitt«. Diese Angaben variieren von Band

51 Z. B. Interviews oder dokumentarische Tonaufzeichnungen.

52 Im Gegensatz zu den meisten elektronischen Medien nahe stehenden elektronischen Editionen, denen die Editionen der Digitalen Bibliothek zuzurechnen sind, hält man hier an einer festen Seitenaufteilung fest.

zu Band. Unter »Konkordanz« wird angegeben, wo in der zugrunde liegenden Buchausgabe die betreffende Textstelle zu finden ist – dies ist besonders wichtig, wenn man die Stelle mit der Vorlage vergleichen will. Die Spalte »Text« gibt an, welches Wort an der betreffenden Stelle genau gefunden wurde. Wie im Beispiel zu sehen, ist dies besonders nützlich, wenn etwa mit Trunkierungen gesucht wurde. Zuletzt wird in den Spalten »Pre-Text« und »Post-Text« noch der vor und nach dem gefundenen Wort stehende Kontext der Fundstelle angezeigt. (Die Tabellendarstellung ist im Beispiel sehr knapp – der Kontext kann auch ausführlicher angezeigt werden.)

Fundstellenlisten können bei einigen elektronischen Editionen auch kommentiert, abgespeichert und somit als Lesezeichen oder Exzerptdatenbanken verwendet werden.



The screenshot shows a search interface with a search bar containing 'greife'. Below the search bar, there are options for 'Filter und Optionen' and 'Suchergebnisse'. The main part of the image is a table with the following columns: 'Seite', 'Akte', 'Buch', 'Abschnitt', 'Konkordanz', 'Text', 'Pre-Text', and 'Post-Text'. The table contains several rows of search results for the word 'greife' in various editions of Goethe's 'Faust: Eine Tragödie'.

Seite	Akte	Buch	Abschnitt	Konkordanz	Text	Pre-Text	Post-Text
50190		Goethe, Johann Wolfgang	Faust: Eine Tragödie	Faust: Der Tragedie zweiter Teil	GoetheFA B.3. 217	greife	I da beherren: Die Sphäre schenkt, umschreibt die Urst was nicht alles, köd'ig und befügt, Vor ihm und
50190		Goethe, Johann Wolfgang	Faust: Eine Tragödie	Faust: Der Tragedie zweiter Teil	GoetheFA B.3. 217	greife	I eigne, Zue oft geschoben, mehr jedoch gelöt, Man nun nach Mätken, Kronen, Gold, Dem Grafenden et
50191		Goethe, Johann Wolfgang	Faust: Eine Tragödie	Faust: Der Tragedie zweiter Teil	GoetheFA B.3. 218	GREIFE	I ausgeat, Sie lachen stot, wie wet sie'n wagger! Un' /'s wollen sie schon zum Genandis bringen, ARMAS,
50201		Goethe, Johann Wolfgang	Faust: Eine Tragödie	Faust: Der Tragedie zweiter Teil	GoetheFA B.3. 226	greife	Ich, Von stächen und die höchste Schatz' gepat, Auf bezüglic, Von desentru, und ohne Feld bewest, No,
50211		Goethe, Johann Wolfgang	Faust: Eine Tragödie	Faust: Der Tragedie zweiter Teil	GoetheFA B.3. 231	GREIFE	sch' katers, 'We lassen uns im heiligen Sit' nicht stören, Gold in Stücken, Gold in Fäden, Durch die Riben seth,
50211		Goethe, Johann Wolfgang	Faust: Eine Tragödie	Faust: Der Tragedie zweiter Teil	GoetheFA B.3. 232	GREIFE	sch' katers, Nur mit dem Gold' heret! Der Berg' laßt fähen, Heret! Heret! Nur Sitk zu Haut' wil legen neue Klas,
50247		Goethe, Johann Wolfgang	Faust: Eine Tragödie	Faust: Der Tragedie zweiter Teil	GoetheFA B.3. 309	greife	auf jedes garstig' Nu, Gelegen'heit' er, du, nun, Faute, auf FAUST, Mit s'dchem Fäbelstern verschone mach' U,
50307		Goethe, Johann Wolfgang	Faust: Eine Tragödie	Faust: Der Tragedie zweiter Teil	GoetheFA B.3. 323	greife	I Ruhel' ist den Gätter' rüht Langer Saal' verkehrt dem Kuzen Wachens' rasches Tun, WUNDER! Sage, Mut,

Abbildung 8. Ausschnitt aus der Fundstellenliste der DIGITALEN BIBLIOTHEK des Verlags Directmedia Publishing

1.4.5 Textexport

Neben den Funktionen, die den Zugriff auf den Textbestand elektronischer Editionen ermöglichen, verfügen einige Editionen über weitere Funktionen, um den Benutzer bei der Arbeit mit den Texten zu unterstützen.

Die Möglichkeit eines direkten Exports von Text aus einer elektronischen Edition in ein anderes Programm, in dem er weiterbearbeitet werden kann, ist neben der *Retrieval*-funktion eine häufig genutzte Funktion elektronischer Editionen. Dazu muss der Text entweder im Volltext vorliegen oder er muss bei einem Image vollständig hinterlegt sein. Beim praktischen Arbeiten mit elektronischen Editionen werden Textstellen häufig als Zitate oder Arbeitstexte in ein Textverarbeitungsprogramm übernommen. Aber auch die Übernahme von Texten in andere Kontexte, etwa eine Website, ist möglich. Dabei sind häufig Urheberrecht und Editionsschutz zu berücksichtigen. Manche elektronischen Editionen unterbinden deshalb einen Export von Text vollständig, um die Interessen des Verlages oder anderer Rechteinhaber zu sichern. In anderen Fällen sind Exportfunktionen eingeschränkt, etwa auf die Menge des Textes, der exportiert werden kann. Nicht immer ist ersichtlich, welche Exporteinschränkungen technisch bedingt sind und welche bewusst eingerichtet wurden. Zur Unterstützung bei der Arbeit mit exportierten Texten fügen manche Editionen automatisch Quellenangaben ein, die oft auch einen Copyright-Hinweis enthalten.

Sehr unterschiedlich fallen die Exportmöglichkeiten verschiedener elektronischer Editionen bezüglich der in der Edition vorkommenden Formatierungen sowie der Übernahme von zugehörigen Fußnoten und Apparaten aus. Einschränkungen sind dabei oft technisch bedingt, da der Austausch von komplexeren Daten, wie sie Formatierungen und verbundene Textteile darstellen, spezielle Datenformate zum Datenaustausch erfordern. Diese können unter Umständen nicht von allen Programmen, die zur Weiterverarbeitung in Frage kommen, verwendet werden.

Eine Möglichkeit wäre, den Text in einer mittels einer *Auszeichnungssprache* strukturierten Form zur Verfügung zu stellen. In der Strukturierung bzw. Auszeichnung elektronischer Texte sehen viele Verlage elektronischer Editionen jedoch ihre spezielle Kompetenz und wahren ihre Interessen, indem sie die strukturierten

Texte nicht zugänglich machen. Teilweise arbeitet die Software der elektronischen Edition auch mit sehr speziellen Auszeichnungen, die nicht für die Verwendung in anderen Kontexten geeignet ist. Positiv hervorzuheben sind hier die elektronischen Editionen von Chadwyck-Healey,⁵³ die in einem einheitlichen in der *Standard Generalized Markup Language* formulierten Standard vorliegen oder die der Edition DER JUNGE GOETHE⁵⁴ beiliegende elektronische Edition, die den Text der Edition unter anderem auch in einer nach den Richtlinien der *Text Encoding Initiative*⁵⁵ kodierten Version enthält.

Neben dem Text besteht bei einigen Editionen auch die Möglichkeit, Fundstellenlisten, Lesezeichen und eigene Kommentare zum Text zu exportieren.

1.4.6 Notizen, Lesezeichen und Exzerptdatenbanken

Als zusätzliche Hilfe bieten einige elektronische Editionen einfache Programme zum Schreiben und Edieren von Text als »Notizblöcke« an. Der Funktionsumfang dieser Notizblöcke fällt zwischen den verschiedenen elektronischen Editionen sehr unterschiedlich aus und reicht von einem schlichten Editor bis zu kompletten Textverarbeitungsprogrammen.

Die Möglichkeit, im Text der Edition Stellen zu markieren und mit eigenen Kommentaren zu versehen, ist eine weitere Funktion, die verschiedene elektronische Editionen bieten. Je nachdem, wie komplex die vorhandenen Möglichkeiten sind, kann sich ein Nutzer daraus eigene Exzerptdatenbanken aufbauen.

1.4.7 Retrodigitalisierung

Unter Retrodigitalisierung oder retrospektiver Digitalisierung versteht man die elektronische Erfassung von Printmedien,⁵⁶ also deren nachträgliche bzw. rückwirkende Umwandlung in elektronische Daten. Je nach angestrebtem Verwendungszweck werden die Vorlagen als elektronische Bilder (images) oder als Zeichendaten erfasst.

Ist das Ziel der Retrodigitalisierung die Erstellung einer elektronischen Image-Edition,⁵⁷ werden die Vorlagen zur Gewinnung von Bilddaten meist gescannt oder digital abfotografiert. Der technische Aufwand, der hierbei betrieben wird, ist sehr unterschiedlich und reicht bei der Verarbeitung großer Textmengen vom Zerschneiden vorhandener Bücher und der Verarbeitung mit Scannern, die die einzelnen Blätter automatisch einziehen (automatischer Einzelblatteinzug) und Vorder- und Rückseite eines Blattes in einem Arbeitsgang erfassen (Duplexscanner) bis zur Konstruktion spezieller »Buchwippen«, die wertvolle Originale beim Abfotografieren möglichst wenig belasten.

Auch bei der Gewinnung von Volltext-Daten steht vielfach die Aufnahme eines Bildes am Anfang. Das Bild wird dann mit Hilfe einer speziellen Software zur automatischen Texterkennung, Optical Character Recognition (OCR, manchmal auch eingedeutscht: Optische Zeichenerkennung OZE) verarbeitet. Die OCR arbeitet mit einer Mustererkennung, d. h. die Software versucht im Bild Muster zu erkennen, die einem Buchstaben entsprechen und ordnet dem erkannten Muster den Buchstaben zu. Eine reine OCR ist jedoch sehr fehleranfällig. Je nach Qualität der gedruckten Vorlage kommt es häufig zu typischen OCR-Fehlern. Beispielsweise gibt es Buchstaben und Zahlen, die ein hohes Fehlerpotential bei der Mustererkennung bieten, etwa »I« und »l« oder »O« und »o« (Null). Auch gibt es kritische Buchstabenfolgen, »rn« (r n) werden etwa oft als »m« erkannt (»Büchem« statt »Büchern«) und umgekehrt. Um diesen Problemen zu begegnen, kann beinahe jede OCR Software auf eine be-

53 Goethes Werke im WWW bzw. Goethes Werke auf CD-ROM; Kafkas-Werke ...; Schillers Werke...; etc.

54 Goethe, Johann Wolfgang von: Der junge Goethe in seiner Zeit. Texte und Kontexte. Hg. von Karl Eibl u. a. Frankfurt a. M.: Insel, 1998. 2 Bände und 1 CD-ROM.

55 Vgl. 3.3 Die Text Encoding Initiative, S. 94.

56 Die Begriffe werden teilweise auch für die Digitalisierung von anderen »analogen« Medien, etwa Filmen oder Tonbändern, verwendet.

57 Vgl. hierzu 1.4.2 Textdarstellung, S. 20.

stimmte Vorlage »trainiert«, d. h. an Schriftart und Schrifteigenheiten der Vorlage angepasst werden. Ergänzend arbeitet man meist noch mit einer automatischen Rechtschreibkontrolle. In diesem Fall spricht man auch von »Intelligent Character Recognition«. In vielen Fällen muss ein Text trotzdem noch manuell Korrektur gelesen werden.



Abbildung 9. Einfacher Desktop-Scanner mit Duplex-Funktion und Einzelblatteinzug



Abbildung 10. Buchscanner mit Buchwippe zur schonenden Auflage von Büchern

Während Antiqua-Schriften meist sehr gut von entsprechenden OCR Programmen erkannt werden, stößt die automatische Texterkennung bei Fraktur-Vorlagen schnell an ihre Grenzen, obwohl es auch auf diesem Gebiet große Fortschritte gibt. Weitere Einschränkungen bei den Möglichkeiten der OCR sind etwa die Qualität der Vorlage oder konservatorische Aspekte. Das Papier der Vorlage kann vergilbt und die Druckfarbe verblasst sein oder der Text kann durch Tintenfraß entstellt sein. Auch kann die mechanische Belastung oder der Lichteinfluss beim Scannen wertvolle Bestände schädigen. Kann eine Vorlage nicht in der nötigen Qualität gescannt, also als elektronisches Bild erfasst werden, scheidet die optische Zeichenerkennung aus.

Können keine Verfahren zur automatischen Texterkennung eingesetzt werden, weil sie keine verwertbaren Ergebnisse liefern, bietet sich die manuelle Erfassung, also das Abtippen der Vorlagen von Menschenhand an. Bei der Fehlerkontrolle kommt häufig das so genannte »doublekeying« zur Anwendung: Die gleiche Vorlage wird unabhängig von zwei Mitarbeitern abgetippt; anschließend werden die Daten elektronisch verglichen. Bei Abweichungen muss ein nochmaliger Vergleich mit der Vorlage erfolgen. Bei sehr hohen Ansprüchen an das Ergebnis wird teilweise sogar ein triplekeying eingesetzt, d. h. drei Personen erfassen die gleiche Vorlage. Anschließend werden die drei Versionen miteinander verglichen. Die automatische Texterkennung per OCR ist meist schneller und preiswerter.

Neben der Qualität der Vorlage hängt es von der geplanten Verwendung ab, wie viel Aufwand zur Herstellung geeigneter Zeichendaten notwendig ist. Sollen die Daten etwa im Rahmen von elektronischen Imageeditionen lediglich den Bilddaten hinterlegt werden, wird meist eine bestimmte Fehlerquote toleriert.⁵⁸

1.5 Hybrideditionen

Neben der Erstellung eigenständiger elektronischer Editionen gibt es auch Versuche, Buchausgaben durch elektronisches Material zu ergänzen. In diesem Fall spricht man häufig von »Hybrideditionen«. Auf einem mitgelieferten Datenträger oder über das Internet wird Material angeboten, das in dieser Form nicht im Buch enthalten ist. Hybridausgaben sind also von Konzepten wie »Buch und CD-ROM«, bei denen dasselbe Material sowohl im Buch als auch in elektronischer Form geboten wird, zu unterscheiden.

Bei den Hybrideditionen, die ich hier kurz vorstelle, wurde jeweils der Versuch unternommen, Daten, die bei der Erstellung der Edition anfielen, für die elektronische Ergänzung zu nutzen.

Das erste Beispiel ist die HISTORISCH-KRITISCHE GOTTFRIED KELLER AUSGABE (HKKA).⁵⁹ Die Bände der Druckausgabe erscheinen sukzessive und werden durch CD-ROMs ergänzt. Zurzeit besteht die elektronische Edition noch aus zwei einander ergänzenden Bereichen, die jedoch im Laufe der Weiterentwicklung unter einer Benutzeroberfläche vereint werden sollen: einem Datenbankauszug der bei der Erstellung der Edition verwendeten Arbeitsdatenbank mit Texten und Materialien sowie einer elektronischen Handschriften-Edition. Die Problematik der Datenbank thematisiert die Projektbeschreibung selber:

Die Datenbank-Edition ist Teil der Datenbanksoftware, welche zur Erfassung und Verwaltung der Editionsdaten entwickelt und später zu Präsentationszwecken weiter ausgebaut wurde. Es handelt sich um ein DOS Programm ohne graphische Oberfläche und ohne Mausunterstützung. Das Programm ermöglicht – neben gebräuchlichen Suchverfahren – die parallele Darstellung von Text und Varianten (bzw. Stellenkommentaren, Querverweisen) und die automatische Erstellung von Textstufen. Da die Handhabung der Datenbank-Edition nicht mehr heutigen Bedürfnissen entspricht, wird das Datenbank-Programm gegenwärtig umgearbeitet und zusammen mit der Handschriften-Edition in eine neue Benutzeroberfläche integriert.⁶⁰

Es ist zu ergänzen, dass das Datenbankprogramm auf aktuellen Computersystemen überhaupt nicht mehr läuft. Im Prinzip besteht die HKKA derzeit also aus drei völlig unterschiedlich aufgearbeiteten und zu benutzenden Komponenten: den Buchbänden, der DOS-Datenbank und der elektronischen Handschriften-Edition. Da die HKKA großzügig mit finanziellen Mitteln ausgestattet wurde, ist die Entwicklung einer besseren, integrativen Lösung in Arbeit. Auf das Ergebnis darf man gespannt sein.

Auch der Edition DER JUNGE GOETHE IN SEINER ZEIT⁶¹ liegt eine CD-ROM bei. Den Kern der elektronischen Edition bildet der Textbestand der Bücher. Dabei wurden jedoch nicht, wie bei dem Konzept »Buch und CD-ROM« sonst häufig, die Satzdaten als elektronisches Dokument im *PDF* beigefügt, vielmehr wurde für die elektronische Edition eine spezielle Software, Folio Views, verwendet. Der Textbestand der Buchausgabe ist als Volltext verfügbar, die Software bietet umfangreiche Recherchemöglichkeiten, ist aber leider nicht sehr komfortabel zu bedienen. Die elektronische Version enthält über den Textbestand der Buchedition hinaus noch umfangreiches Kontextmaterial, das allerdings in einer sehr heterogenen Form vorliegt: teils im Volltext, teils in Form von Images. Trotz der recht komplexen Software ist die Verbindung zwischen der eigentlichen Edition und dem ergänzenden Material nicht gut gelungen und das Editions-konzept nicht schlüssig. Hervorzuheben ist, dass auf der CD-ROM zusätzlich der Textbestand der Bücher in einer mittels *Standard Generalized Markup*

58 Vgl. etwa Adelung, Johann Christoph: Grammatisch-kritisches Wörterbuch in der elektronischen Edition der Bayerischen Staatsbibliothek: <http://mdz.bib-bvb.de/digbib/lexika/adelung> (18.05.2008).

59 Keller, Gottfried. Sämtliche Werke. Hg. von Walter Morgenthaler. Basel u. a.: Stroemfeld u. a., 1996 ff.

60 <http://www.gottfriedkeller.ch/hkka/hkka.htm> (18.05.2008).

61 Goethe, Der junge Goethe in seiner Zeit.

Language nach den Richtlinien der *Text Encoding Initiative (TEI P3)* ausgezeichneten Version vorliegt. Ohne weitere Hilfsmittel (wie etwa einem Browser oder auch nur geeigneten *Stylesheets* oder *Templates*) kann man mit den Daten aber wenig anfangen. Immerhin wird somit der erarbeitete Textbestand auch unabhängig von der ansonsten verwendeten Software verfügbar gehalten.

1.6 »Status quo« elektronischer Editionen

Die meisten der zurzeit verfügbaren elektronischen Editionen lassen sich den anfangs beschriebenen Typen »buchnah« oder »elektronischen Medien nah« zuordnen. Die Gründe dafür sind zum einen praktischer Art. Wie ich aufgezeigt habe, bieten sich buchnahe elektronische Editionen als »zweite Verwertungsschiene« bei neuen Bucheditionen an, während elektronischen Medien nahe elektronische Editionen meist auf der Grundlage vorhandener Buchausgaben entstehen. Zum anderen sind die Gründe konzeptueller Art. Die jeweiligen Herausgeber haben sich bewusst für »buchnahe« oder »elektronischen Medien nahe« elektronische Editionen entschieden.

Wie ich aufgezeigt habe, werden buchnahe elektronische Editionen meist parallel zum Buchsatz erstellt. Bei ihrer Erstellung überwiegt die Vorstellung, in erster Linie für die Publikation auf Papier zu arbeiten. Das »papierne Denken« bestimmt im Wesentlichen die Arbeit an der Edition. Erst zu einem im Projektablauf recht späten Zeitpunkt, zu dem die wichtigsten editorischen Entscheidungen (vor dem Hintergrund, in erster Linie eine Buchedition zu erstellen) gefallen sind, wird eine elektronische Edition in die Überlegungen einbezogen.

Viele der elektronischen Medien nahe stehenden elektronischen Editionen beruhen dagegen auf älteren, abgeschlossenen Buchausgaben. Bei abgeschlossenen bzw. bereits bewährten Editionen ist die Materiallage geklärt. Der edierte Text selber liegt bereits nach den Maßgaben der angewandten editorischen Prinzipien in Form gedruckter Bücher vor. Die Umsetzung in elektronische Medien stellt einen eigenen Arbeitsprozess dar. Eine solche »nachträgliche« elektronische Edition muss ihre Eigenständigkeit gegenüber dem Buch behaupten, unterliegt jedoch nicht mehr der Verpflichtung, die maßgebliche Edition zu sein – diese liegt ja bereits in gedruckter Form vor. Ihre Eigenständigkeit bzw. ihren Mehrwert gegenüber der Buchausgabe rechtfertigen sie über ihre elektronischen Funktionen.

Ein weiterer Grund dafür, dass viele den elektronischen Medien nahe stehende elektronischen Editionen auf gedruckten Ausgaben beruhen, ist sicher auch, dass die Deutsche Forschungsgemeinschaft (DFG) solche Projekte durch den »Förderbereich zum Aufbau einer ›Verteilten Digitalen Forschungsbibliothek« und das Förderprogramm »Retrospektive Digitalisierung von Bibliotheksbeständen« unterstützt hat.⁶²

1.6.1 Verfügbarkeit elektronischer Editionen

Elektronische Editionen sind im Vergleich zu Bucheditionen oft recht kurzlebig. Der Grund hierfür liegt in der ständigen Weiterentwicklung von Computern und Software. Im Gegensatz zum Buch, das einmal gedruckt über Jahrhunderte hinweg benutzt wird, benötigen die meisten elektronischen Editionen eine spezielle Software, mit deren Hilfe auf die Texte zugegriffen werden kann. Wird eine solche Software nicht gepflegt, d. h. wird sie nicht an neue Computersysteme angepasst und weiterentwickelt, ist sie im Laufe weniger Rechnergenerationen nicht mehr zu benutzen.⁶³ Besonders bei Verlagen, die solche Editionen einmalig herausgeben, wird eine Pflege der Software selten gewährleistet. Die meisten Verlage betrachten eine Edition mit ihrem Erscheinen als abgeschlossen. Dieser Umstand wird durch die Schnelllebigkeit von Firmen, die sich mit elektronischen Medien be-

62 Seit dem Abschluss des Förderprogramms sind die diesbezüglichen Informationen der DFG leider nicht mehr verfügbar. Eine Liste der geförderten Projekte findet sich unter <http://www.dl-forum.de> (18.05.2008). Vgl. auch Altrichter, Helmut: Retrodigitalisierung in Deutschland – Versuch einer Zwischenbilanz. <http://www.bsb-muenchen.de/mdz/forum/altrichter/> (18.05.2008).

63 Welche Aspekte für die Erstellung zukunftssicherer elektronischer Editionen zu berücksichtigen sind, erläutere ich unter 2.3.2 Erarbeitung zukunftssicherer elektronischer Editionen, S. 67.

schäftigen, verstärkt. Ein prominentes Beispiel ist etwa *KINDLERS NEUES LITERATURLEXIKON AUF CD-ROM*.⁶⁴ Ursprünglich von der Firma Systema herausgegeben, wurde es von verschiedenen anderen Firmen übernommen, über unterschiedliche Wege vermarktet und ist mittlerweile kaum noch zu bibliografieren.

Ein positives Gegenbeispiel bildet der *INDEX THOMISTICUS*, mit dessen Erstellung Roberto Busa 1949 in Italien begann. Das Ergebnis wurde zuerst als 56-bändiges Werk in Buchform veröffentlicht und ist in elektronischer Form bis heute im Internet verfügbar.⁶⁵

Das Altern elektronischer Datenträger, also ihre Archivbeständigkeit, ist in diesem Zusammenhang ein weiterer, viel diskutierter Punkt, auf den ich später eingehen werde.

Daneben gibt es häufig auch Probleme bei der Nutzung elektronischer Editionen, wenn diese in Bibliotheken zur Verfügung gestellt werden. An der Staatsbibliothek zu Berlin gibt es Probleme, da einige Editionen nicht an allen zur Verfügung gestellten Rechnern zu benutzen sind, weil sie nicht die von der Software geforderte Bildschirmauflösung besitzen; bei anderen Editionen arbeiten einzelne Funktionen nicht, da diese andere Systemvoraussetzungen benötigen. Manchmal herrscht auch beim Bibliothekspersonal schlichte Unkenntnis beim Umgang mit elektronischen Medien. Als ich in der philologischen Bibliothek der Freien Universität Berlin die CD-ROM zu »Der Junge Goethe und seine Zeit« einsehen wollte, führte man mich in eine Audio-Kabine.

1.6.2 Tabellarischer Überblick

Da elektronische Editionen nicht so verbreitet sind, dass ich davon ausgehen kann, dass sie jedem in repräsentativer Auswahl zugänglich sind, und um dem Umstand Rechnung zu tragen, dass sicher einige der für diese Arbeit herangezogenen Editionen der schnellen Vergänglichkeit elektronischer Medien und ihren Inhalten anheim fallen, habe ich eine Überblickstabelle verbreiteter elektronischer Editionen zusammengestellt. Sie erhebt keinen Anspruch auf Vollständigkeit. Bei Editionen, die einem Konzept folgen bzw. denen eine bestimmte Software zugrunde liegt, etwa den Editionen der *DIGITALEN BIBLIOTHEK*⁶⁶ oder von Chadwyck-Healey⁶⁷, habe ich mich auf einzelne Editionen beschränkt.

64 Kindlers neues Literatur-Lexikon. Das 23-bändige Werk auf einer CD-ROM. München: Systema u. a., 2000.

65 Busa, Roberto: *Index Thomisticus. Sancti Thomae Aquinatis operum omnium indices et concordantiae in quibus verborum omnium et singulorum formae et lemmata cum suis frequentis et contextibus variis modis referuntur, quaeque auspice Paulo VI Summo Pontifice consociata plurium opera atque electronico IBM automato usus digessit Robertus Busa.* Stuttgart, Bad Cannstadt: Frommann-Holzboog, 1974–80. Im Internet unter <http://www.corpusthomisticum.org> (18.05.2008).

66 In der Reihe der Digitalen Bibliothek des Verlags Directmedia sind über 150 elektronische Editionen unterschiedlicher Inhalte erschienen. Die Reihe wird jedoch eingestellt. Einzelne Titel werden unter dem Label Zeno.org weitergeführt.

67 Die Firma Chadwyck-Healey ist der international größte Anbieter elektronischer Texte im geisteswissenschaftlichen Bereich. Die meisten Editionen erschließen englischsprachige Literatur, in Zusammenarbeit mit verschiedenen deutschen Verlagen sind auch einige deutsche Volltextdatenbanken herausgegeben worden.

	Medium	Software	Editionsart	Konkordanz	Darstellung		Navigation		
					skalierbar	Seitenumbrüche	Inhaltsbaum	Register	Hypertext
Deutsche Literatur von Lessing bis Kafka	C	eig.	V	•	•	fest	•	•	•
Projekt Gutenberg-DE	N	B	V	—	•	—	—	•	•
Der Junge Goethe in seiner Zeit	C	Fv	V ²	•	•	—	•	•	•
Goethes Werk im Kontext	C	eig.	V	•	•	fest	•	•	•
Goethes Werke auf CD-ROM/im WWW	C/N	eig./B	V	•	•	—	•	•	•
Johann Wolfgang von Goethe. Werke	C	eig.	V	•	•	—	•	•	•
Johann Wolfgang von Goethe. Zeit, Leben, Werk	C	eig.	V	—	—	—	—	•	•
Heinrich-Heine-Portal)	I	B	V	•	•	fest	•	—	•
Franz Kafka: Historisch-kritische Ausgabe	C	Ar	V/I	•	—	fest	•	—	—
Kafkas Werke auf CD-ROM/im WWW	C/N	eig./B ³	V	•	•	—	•	•	•
Historisch-Kritische Gottfried Keller Ausgabe	C	eig.	V/I	•	•	fest	—	•	•
Deutsches Wörterbuch	C/N	eig./B	V	•	•	—	—	•	—
Mittelhochdeutsche Wörterbücher	C/N	eig./B	V	•	•	—	—	•	•
Adelung, Wörterbuch	N	B	V/I	— ⁵	•/—	•/—	•	—	—
J. Chr. Adelung: Grammatisch-kritisches...	C	eig.	V/I	•	•	fest	•	•	•
KLG	C/N	eig./B.	V	—	•	•	—	•	—
Kindlers neues Literaturlexikon	C	eig.	V	—	•	—	—	•	•
Die Fackel	C	eig.	V	•	•	fest	•	•	—
Exilpresse digital	N	B	I	•	—	fest	—	•	—
Rezensionsorgane und Literaturzeitschriften.	N	B	I	•	—	fest	—	•	—

C Containermedium (CD-ROM, DVD)

N Netzwerkressource (Internet)

V Volltextedition

I Image-Edition

Fv FolioView

Ar Acrobat Reader

Db Datenbank

Hb Harvest Brooker

1 Funktion des Internetbrowsers

2 ergänzendes Material teilw. als Image

3 Abhängig vom Medium, in der CD-ROM Version umfangreicher

4 Funktionen der Acrobat-Software

Retrieval														
Volltext	Suchmöglichkeiten						Ergebnisanzeige						Export	Technik
	eingeschränkt	tolerant	Trunkierungen	boolesch	speichern	Hervorhebung	Liste	Kontext	speichern	Leseseichen	Kommentare			
•	•	•	•	•	•	•	•	•	•	•	•	•	•	XML
• ⁵	–	–	–	–	–	• ⁵	–	–	–	–	–	–	•	HTML
•	•	•	•	•	•	•	•	•	•	•	•	•	•	Fv
•	•	•	•	•	•	•	•	•	•	•	•	•	•	Word
•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	SGML
•	•	•	•	•	•	•	•	•	•	•	•	•	•	XML
•	•	–	–	–	–	•	–	–	–	–	–	–	–	HTML
•	•	–	–	•	–	•	•	•	–	–	–	–	–	XML
•	–	–	• ⁴	–	–	•	•	–	–	• ⁴	• ⁴	–	–	PDF
•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	•/– ³	SGML
•	•	–	–	–	–	•	–	–	–	–	–	–	–	Db
•	•/– ³	–	•/– ³	•/– ³	•/– ³	•	•	–	•/– ³	–	–	–	•	SGML
•	•/– ³	•/– ³	•/– ³	•/– ³	–	–	•	–	–	•	•	•	•	SGML
•	–	•	•	•	–	•	•	–	–	–	–	–	•	?
•	•	•	•	•	–	•	•	•	•	•	•	•	•	XML
•	•	•	•	•	•	•	•	•	•	•	•	•	•	?
•	•	–	•	•	–	•	•	–	–	•	•	•	•	Db
•	•	–	•	•	–	•	•	•	–	–	–	–	–	?
•	–	•	•	–	–	–	•	–	–	–	–	–	–	Hb
–	–	–	–	–	–	–	•	–	–	–	–	–	–	SGML ⁶

5 nur über Image

6 SGML zur Erschließung bibliographischer Daten genutzt

- Bertram, Mathias (Hg.): Deutsche Literatur von Lessing bis Kafka. Studienbibliothek. Berlin: Directmedia Publishing, 2000.
- Projekt Gutenberg-DE: <http://gutenberg.spiegel.de> (18.05.2008)
- Der junge Goethe in seiner Zeit. Texte und Kontexte. Hg. von Karl Eibl u. a. Frankfurt a. M.: Insel, 1998. 2 Bde. und 1 CD-ROM.
- Goethes Werk im Kontext. Die »Berliner Ausgabe« auf CD-ROM. Berlin: Carsten Worm InfoSoftWare, 2004.
- Goethes Werke auf CD-ROM/Goethes Werke im WWW. Weimarer Ausgabe. Cambridge: ProQuest Information and Learning Co., 1997-
<http://www.goethe.chadwyck.com> (18.05.2008).
- Johann Wolfgang von Goethe. Werke. Hg. von Mathias Bertram. CD-ROM. Berlin: Directmedia 1998.
- Johann Wolfgang von Goethe. Zeit, Leben, Werk. Von Jürgen von Esenwein/Harald Gerlach. Stuttgart u. a.: Metzler u. a. 1999.
- Heinrich Heine Portal: <http://germazope.uni-trier.de/Projects/HHP/> (18.05.2008).
- Franz Kafka. Historisch-kritische Ausgabe sämtlicher Handschriften, Drucke und Typoskripte. Hg. von Roland Reuß. Basel/Frankfurt a. M.: Stroemfeld, 1995 ff.
- Kafkas Werke auf CD-ROM/Kafkas Werke im WWW. Kritische Kafka-Ausgabe des S. Fischer Verlages bei Chadwyck-Healey. Chadwyck-Healey, 1999.
<http://kafka.chadwyck.com> (18.05.2008).
- Gottfried Keller. Sämtliche Werke. Hg. von Walter Morgenthaler. Basel u. a.: Stroemfeld u. a., 1996 ff. <http://www.gottfriedkeller.ch/hkka/hkka.htm> (18.05.2008).
- Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm. Elektronische Ausgabe der Erstbearbeitung. Hg. vom Kompetenzzentrum für Elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier... Bearb. von Hans-Werner Bartz... Frankfurt a. M.: Zweitausendeins, 2004. <http://germazope.uni-trier.de/Projects/DWB> (18.05.2008).
- Mittelhochdeutsche Wörterbücher im Verbund. CD-ROM und Begleitbuch. Hg. von Thomas Burch, Johannes Fournier und Kurt Gärtner. Stuttgart: Hirzel, 2002. <http://gaer27.uni-trier.de/MWV-online/MWV-online.html> (18.05.2008).
- Adelung, Johann Christoph: Grammatisch-kritisches Wörterbuch der Hochdeutschen Mundart: mit beständiger Vergleichung der übrigen Mundarten, besonders aber der Oberdeutschen. Mit D. W. Soltau's Beyträgen revidirt und berichtigt von Franz Xaver Schönberger. Wien: Bauer, 1811.
<http://mdz.bib-bvb.de/digbib/lexika/adelung> (18.05.2008).
- Adelung, Johann Christoph: Grammatisch-kritisches Wörterbuch der hochdeutschen Mundart: mit beständiger Vergleichung der übrigen Mundarten, besonders aber der Oberdeutschen. Elektronische Volltext- und Faks.-Ed. nach der Ausg. letzter Hand, Leipzig 1793–1801. Berlin: Directmedia Publishing, 2001.
- KLg – Kritisches Lexikon zur deutschsprachigen Gegenwartsliteratur. Hg. von Heinz Ludwig Arnold. München: Edition Text Kritik. Als Loseblattsammlung, auf CD-ROM und online. <http://www.klgonline.de> (18.05.2008).
- Kindlers neues Literatur-Lexikon. Das 23-bändige Werk auf einer CD-ROM. München: Systema u. a., 2000.
- Die Fackel. Volltextausgabe. Hg. von Karl Kraus. Mit Bibliographie und Register von Wolfgang Hink. München: Saur; Frankfurt a. M.: Zweitausendeins, 2002.
- Exilpresse Digital. Deutsche Exilzeitschriften 1933–1945: <http://deposit.ddb.de/online/exil/exil.htm> (18.05.2008).
- Rezensionsorgane und Literaturzeitschriften: Retrospektive Digitalisierung wissenschaftlicher Rezensionsorgane und Literaturzeitschriften des 18. und 19. Jahrhunderts aus dem deutschen Sprachraum. <http://www.ub.uni-bielefeld.de/diglib/aufklaerung/index.htm> (18.05.2008).

2 Text in Büchern und elektronischen Medien

Elektronischen Medien nahe Editionen haben, wie ich in Kapitel 1.2 aufgezeigt habe, meist eine bestehende, gedruckte Edition als Referenz. Dagegen entstehen dem Buch nahe stehende elektronische Editionen oft im Zusammenhang mit einer neuen Buchedition.

Für elektronischen Medien nahe stehende Editionen, die auf bestehenden und etablierten Buchausgaben beruhen, muss die Buchvorlage im allgemeinen retrodigitalisiert und für elektronische Medien eingerichtet werden. Dabei steht die Einrichtung des Textes für die Nutzung in elektronischen Medien im Vordergrund, da die elektronische Edition einen Mehrwert gegenüber der bereits vorhandenen Buchedition bieten soll, denn warum sollte jemand eine elektronische Edition benutzen, wenn er mit gleichem Nutzen zur vertrauten Buchausgabe greifen kann? Entsprechend stehen bei solchen Editionen die speziellen Möglichkeiten des elektronischen Zugriffs auf den Textbestand im Vordergrund. Das Ausnutzen der technologischen Möglichkeiten gehört hier zum Konzept der elektronischen Edition. Die als Grundlage dienende Buchedition »entlastet« dabei die elektronische Edition, da sie als Referenz zur Verfügung steht. So kann man beispielsweise die Vorteile der elektronischen Edition beim Auffinden von Textstellen nutzen, kann sich aber beim Zitieren auf die vertraute Buchausgabe berufen.

Im Falle buchnaher elektronischer Editionen ist meist das zu erstellende Buch das Referenzprodukt, die elektronische Edition hat den Charakter eines Nebenproduktes, das sich stark am Buch orientiert. Solche elektronischen Editionen werden gern in Verbindung mit einer Buchausgabe herausgegeben. Die elektronische Edition ist dabei eine Beigabe zum gedruckten Buch und wird als »Buch mit CD-ROM« vertrieben.

Dass elektronische Editionen generell einen starken Bezug zu Buchausgaben haben, ist nicht verwunderlich. Vergleicht man, wie lange Texte schriftlich in Büchern bzw. deren Vorformen – Tontafeln, Papyrusrollen, Kodizes⁶⁸ – überliefert werden und wie kurz im Vergleich dazu der Zeitraum ist, in dem es elektronische Editionen oder überhaupt Computer gibt, wird klar, dass Bücher eine ganz andere Tradition haben: Texte werden seit mehr als 5000 Jahren mit Hilfe von Schrift überliefert. Papier und Bücher bzw. Kodizes gibt es seit über 2000 Jahren. Vor 500 Jahren »erfand« Gutenberg den Druck mit beweglichen Bleiletern.

Entsprechend lange sind Texte an Schrift und Schrifträger gebunden. Sprache mittels Schrift auf einem Trägermaterial zu fixieren war lange Zeit eine Schlüsseltechnologie verschiedener Kulturen. Durch den Buchdruck haben sich in unserem Kulturraum Text, Schrift und Buch und gesellschaftliche Entwicklung wechselseitig beeinflusst und zu einer Tradition bzw. einer Kultur geführt, die retrospektiv gern als »Gutenberg-Galaxis«⁶⁹ bezeichnet wird. Der Begriff verweist auf die zentrale Bedeutung, die Bücher als Medien zur die Speicherung und Weitergabe von Informationen in unserer Kultur gewonnen haben, und zeigt, wie sehr Bücher unser Weltbild geprägt haben.

Computer wurden dagegen erst vor etwa 70 Jahren entwickelt.⁷⁰ Die CD-ROM als erstes Speichermedium, das Kapazitäten für größere Datenmengen, etwa Texte und Bilder, bot, kam 1979 auf den Markt. Das Internet auf Grundlage der heute verwendeten Protokolle steht seit 1982 zur Verfügung. Elektronische Medien können im Vergleich zu Büchern nur auf eine sehr kurze Tradition zurückblicken.

68 Vgl. Assmann, Aleida: Aspekte einer Materialgeschichte des Lesens. In: Gestern begann die Zukunft. Entwicklung der gesellschaftlichen Bedeutung der Medienvielfalt. Hg. von Hilmar Hoffmann. Darmstadt: WBG, 1994. S. 7 ff.

69 Der Begriff der Gutenberg-Galaxis geht weit über die Aspekte von Sprache, Text, Schrift und Buch hinaus und versucht den Stellenwert gedruckter Texte in den westlichen Gesellschaften zu verdeutlichen. Er wurde 1962 von Marshall McLuhan durch sein gleichnamiges Buch geprägt und wird seitdem besonders von Medienwissenschaftlern aufgegriffen. Vgl. McLuhan, Marshall: The Gutenberg Galaxy: The Making of Typographic Man. London: Routledge & Paul, 1962. Deutsch: Die Gutenberg-Galaxis. Das Ende d. Buchzeitalters. Düsseldorf u. a.: Econ Verl., 1968.

70 In den Jahren 1937/38 wurden von verschiedenen Seiten die ersten Konzepte frei programmierbarer Rechner vorgestellt: 1937, Konrad Zuse lässt Elemente der »Von-Neumann-Architektur« patentieren; John Atanasoff baut einen der ersten Digitalrechner, den »Atanasoff-Berry-Computer«; Alan Turing beschreibt die »Turing-Maschine«. 1938 stellt Konrad Zuse die Zuse Z1 fertig.

Der Rückgriff auf das etablierte Buch als Referenz für Texteditionen in elektronischen Medien ist damit naheliegend und verständlich. Der Umgang mit elektronischen Textbeständen ist für die meisten Computernutzer einfacher, wenn Computer Papier simulieren, wenn die Benutzeroberfläche den von Büchern her gewohnten Umgang mit Texten bietet. Dadurch, dass der Computer quasi Papier simuliert und uns in vielen Belangen eine vom Papier her gewohnte Arbeitsweise ermöglicht, gerät die digitale Technik elektronischer Medien, die anderen Regeln unterliegt als die Medien der Gutenberg-Galaxis, aus dem Blickfeld. Es ist demnach notwendig zu untersuchen, inwieweit Text und Texttechnologien mit der Tradition von Büchern verbunden sind und welche Probleme sich bei ihrer Übertragung in ein anderes Medium ergeben. Im Rahmen dieser Arbeit beschränke ich mich auf die Darstellung einiger weniger Aspekte, die vor dem Hintergrund der Tradition von Büchern evident erscheinen, die aber in Bezug auf Texte in elektronischen Medien reflektiert werden müssen.

Dass sich ein neues Medium immer erst im Vergleich zum traditionellen Medium etablieren muss, erscheint naheliegend:

Als natürlich galt und gilt immer das angestammte System, die gewohnten Medien und die überkommenen Kodierungen der Informationen. Neue Systeme, Medien und Kodierungsformen erscheinen dem gegenüber zunächst als artifiziell – so lange, bis sich die Menschen an sie gewöhnt haben. Spätestens dann, wenn wiederum neue Medien auftauchen, wird ihnen in der Opposition die Position der Natürlichkeit zugeschrieben.⁷¹

Mein Ziel ist es, im Folgenden zunächst aufzuzeigen, dass sich im Rahmen der Buchkultur Konventionen bei der Tradierung von Texten entwickelt haben, die mittlerweile so allgemein verbreitet und akzeptiert sind, dass sie selten reflektiert werden. Neben den Informationen, die ein geschriebener Text explizit in Form der Buchstaben enthält, ergeben sich durch die Verschriftlichung von Texten weitere Aspekte, die auf unterschiedlichen Ebenen Informationen enthalten und vermitteln und die somit in Bezug auf die Bedeutung und den Informationsgehalt von Texten relevant sind. Solche Aspekte sind beispielsweise die Anordnung und Gestaltung der Buchstaben auf dem Papier oder die Anordnung und Aufteilung von Texten in Büchern. Im Fall eines Medienwechsels, wie er beim Übergang vom Buch zu elektronischen Medien vorliegt, werden die Konventionen des Mediums Buch deutlich, nämlich genau dann, wenn Informationen, deren Übermittlung an solche Konventionen gebunden sind, im anderen Medium verloren gehen.

2.1 Texte in Büchern

Wenn von »Texten in Büchern« die Rede ist, umfasst dies mehrere Aspekte.

- Es handelt sich um einen Text, der mittels Schrift auf Papier geschrieben oder gedruckt ist, wobei das Papier in einer bestimmten Form, als Buch, geordnet und gebunden ist.
- Die gedruckten Texte sind gestaltet. Dabei haben Form und Anordnung der Buchstaben und Absätze (Typografie und Layout) neben der ästhetischen auch eine semantische Dimension.
- Zur Verbreitung von Büchern haben sich bestimmte Distributionswege etabliert. Organisation und Struktur des Buchhandels oder des Bibliothekswesens haben ebenfalls einen Einfluss auf die Bewertung und Interpretation der über sie verbreiteten Texte.

Der Begriff »Text« wurde aus verschiedenen Perspektiven diskutiert und hinterfragt. Ohne auf die »schwierige und ganze Bibliotheken füllende linguistische Diskussion um den Textbegriff an sich«⁷² einzugehen, greife ich

71 Giesecke, Michael: *Sinnenwandel, Sprachwandel, Kulturwandel. Studien zur Vorgeschichte der Informationsgesellschaft*. Frankfurt a. M.: Suhrkamp, 1992. S. 36 f.

72 Gronemeyer, Horst: *Der Philologe und sein Text in Handschrift, Buch und Datenbank: Vorgelegt in der Sitzung vom 18. 1. 2002*. Göttingen: Vandenhoeck & Ruprecht u. a., 2002. S. 4.

unter grober Vereinfachung der komplexen Entwicklung von Schriftsystemen und ihren Rückwirkungen auf die gesprochene Sprache einige Aspekte heraus, die mit dem Aufschreiben von Texten einhergingen. Dabei möchte ich den Begriff »Text« nicht auf eine Definition wie »schriftlich auf Papier fixierte Informationen« einschränken, geht es in dieser Arbeit doch im Wesentlichen um Texte in elektronischen Medien und darum, die Bindung von Texten an Papier zu thematisieren. Die Suche nach einer konkreten Definition des Textbegriffes scheint für Philologen, Literaturwissenschaftler oder Editionswissenschaftler auch nicht zwingend notwendig zu sein.⁷³ »Philologen leben [...] relativ bequem mit einem Grundbegriff, der sich offensichtlich einer präzisen Definition entzieht.«⁷⁴

Das »Schreiben« steht im Blickpunkt verschiedener wissenschaftlicher Disziplinen. Medientheorie, literaturwissenschaftliche Schreibprozessforschung, Textkritik, Diskursgeschichte, aber auch Kognitionspsychologie und Linguistik nähern sich dem Begriff bzw. der dahinterstehenden Tätigkeit mit unterschiedlichen Interessen und unter verschiedenen Gesichtspunkten. Die Bedeutung der »Schrift« selber und damit der Typografie ist in den Geisteswissenschaften bisher selten thematisiert worden:

In der Literaturwissenschaft ist nicht nur das Bewußtsein kaum ausgeprägt, es fehlen auch Konzepte, die Typographie als nicht hintergehbare Bedeutungsdimension (nicht nur literarischer) Texte zu thematisieren und den Einsatz von Typographie auch jenseits der literarischen Avantgarden zu beschreiben und zu analysieren erlauben.⁷⁵

Schließlich lohnt sich ein Blick auf die Bedeutung der Erfindung des Buchdrucks und der damit verbundenen Institutionalisierung des Buchwesens mit seinen Produktions- und Distributionsaspekten.

2.1.1 Geschichte als Heuristik

Die meisten Leser werden diese Arbeit in »Papierform« lesen: als Ausdruck der Computerdatei, die ich mit Hilfe eines Computers und eines Textverarbeitungsprogrammes verfasse, oder gar als gedrucktes Buch. Ich schreibe sie vor dem Hintergrund eines noch immer stark dem Buch als Informationsträger verhafteten Lehrbetriebs. Auch wenn ich elektronische Medien als Hilfsmittel hinzuziehe, liegen dem geisteswissenschaftlichen Teil dieser Arbeit vornehmlich Bücher zugrunde. Damit ist diese Arbeit unweigerlich dem »papiernen Denken« verbunden. Aus dieser Position heraus ist es schwierig, die medialen Aspekte schriftlicher Texte zu reflektieren.

Da die Eigenheiten eines Mediums immer dann am stärksten hervortreten, wenn das Medium neu auftaucht und sich gegen »angestammte Systeme« behaupten muss,⁷⁶ scheint mir ein geschichtlicher Rückblick auf die Entwicklung des Buches eine geeignete Heuristik zu sein, mich den genannten Aspekten von Texten in Büchern zu nähern. Als das Buch neu war, musste es sich ebenfalls gegen die mündliche Tradierung von Texten als »angestammtes System« durchsetzen, wobei seine Eigenheiten hervortraten.

Auf den ersten Blick wäre es naheliegend, einen geschichtlichen Rückblick im 15. Jahrhundert zu beginnen. 1452 bis 1454 entstand in Mainz die 42-zeilige Gutenbergbibel (B42). Damit begann (in unserer abendländischen Kultur) das Zeitalter des Buchdrucks. Die wichtigsten Entwicklungen auf dem Weg zu unserem heutigen Verständnis und Umgang mit schriftlichen Texten lassen sich jedoch ins 12. Jahrhundert zurückverfolgen.

Schriftliche Überlieferungen gibt es seit über 6000 Jahren. Die in unserem Kulturraum auch heute noch maßgeblichen Techniken, um Texte in Büchern zu tradieren, entstanden oder etablierten sich jedoch im

73 Vgl. stellvertretend für die Diskussion um den Textbegriff in der Literaturwissenschaft: Baßler, Moritz: Stichwort Text. Die Literaturwissenschaft unterwegs zu ihrem Gegenstand. In: Jahrbuch der Deutschen Schiller-Gesellschaft 42 (1998). S. 470-475.

74 Wallmannsberger, Josef: Virtuelle Textwelten. Theoretische Grundlagen und Implementationsperspektiven der anglistischen Computerphilologie. Heidelberg: Winter, 1994. S. 30.

75 Wehde, Susanne: Typographische Kultur. Eine zeichentheoretische und kulturgeschichtliche Studie zur Typographie und ihrer Entwicklung. Tübingen: Niemeyer, 2000. S. 21.

76 Vgl. Giesecke, Michael: Sinnewandel, Sprachwandel, Kulturwandel. Studien zur Vorgeschichte der Informationsgesellschaft. Frankfurt a. M.: Suhrkamp, 1992. S. 36.

Mittelalter.⁷⁷ Als wichtigster Schritt auf dem Weg zum »Schriftbild der Moderne« wird dabei allgemein die Aufgabe des Prinzips kontinuierlichen Schreibens, der »scriptio continua«, zugunsten der Trennung einzelner Wörter mittels Spatien (Leerzeichen) angesehen, die sich seit dem 8. Jahrhundert von Irland aus im europäischen Raum durchsetzte. Dazu kamen weitere Techniken, die meist nicht neu erfunden wurden, die sich in diesem Zeitraum aber auf breiter Ebene durchzusetzen begannen. Zu nennen ist etwa die Gliederung schriftlicher Texte durch Rubriken, die über eine Art Inhaltsverzeichnis zugänglich gemacht wurden, eine Technik, die sich schon in römischen Stadtgesetzen des ersten nachchristlichen Jahrhunderts findet.⁷⁸ Daneben gab es, etwa bei Hieronymus, die Technik, Sinnzusammenhänge bzw. Satzglieder zeilen- oder absatzweise *per cola et commata* aufzuschreiben.⁷⁹

Man kann demnach nicht sagen, dass wesentliche Techniken unseres modernen Textverständnisses im Mittelalter »erfunden« wurden.

Von daher dürfte es zutreffender sein, die Zeit ab dem 11./12. Jh. [...] als eine Phase von Verschriftlichungsschüben zu betrachten, in der bei Auslaufen frühmittelalterlicher Traditionen die Basis für die moderne europäische Schriftkultur gelegt wurde.⁸⁰

Die Veränderungen im Umgang mit schriftlichen Texten sind vielmehr vor dem Hintergrund des gesellschaftlichen Wandels in dieser Zeit zu betrachten: Mit der zunehmenden Bedeutung der Städte und den damit verbundenen soziokulturellen Veränderungen verloren herkömmliche Bildungseinrichtungen wie Klöster ihre Bedeutung zugunsten der aufkommenden Stadtschulen, aus denen sich die Universitäten mit ihrem Lehrbetrieb entwickelten. Im Umfeld des universitären Lehrbetriebs kam dem Lesen und Schreiben eine neue Bedeutung zu.⁸¹ Bücher dienten nicht mehr hauptsächlich dem meditativ-murmelnden (Vor-)Lesen, sondern wurden, etwa von Schülern der Universitäten, still gelesen.⁸² Die Rolle des Autors änderte sich. Anstatt Texte zu diktieren, kam es immer mehr dazu, dass Autoren Texte selber schrieben. Als einer der ersten »Selbstschreiber« des Mittelalters gilt z. B. Thomas von Aquin.⁸³ Schüler lasen Texte zur Vorbereitung auf eine Lehrveranstaltung und schrieben bei den Vorlesungen ihrer Magister mit. Dazu kamen weitere Veränderungen. Durch die Entwicklung eines komplexeren Handelwesens gewann die Fähigkeit, schreiben und lesen zu können, für Kaufleute eine größere Bedeutung. Durch die zunehmende Schreibkompetenz außerhalb kirchlicher Einrichtungen entstanden neue Textgattungen. Zum einen wurden andere Sprachen als das in der Kirche vorherrschende Latein geschrieben, zum anderen wurden außerhalb des scholastischen Lehrbetriebes auch andere Texte bzw. Textsorten aufgeschrieben.

Vor diesem Hintergrund entwickelte sich ein neuer Umgang mit Texten, der neue Techniken beim Schreiben und Lesen etablierte.

Durch diese Ansammlung von Techniken und Gewohnheiten wurde es möglich, sich den Text als etwas von der physischen Realität der Buchseite Losgelöstes vorzustellen. Sie reflektierte und revolutionierte zugleich, was gebildete Menschen taten, wenn sie lasen – und als was sie das Lesen erlebten.⁸⁴

77 Vgl. Illich, Ivan: Im Weinberg des Textes. Als das Schriftbild der Moderne entstand. Ein Kommentar zu Hugos »Didascalicon«. Frankfurt a. M.: Luchterhand, 1991. Illich datiert den Beginn des »neuen Buches« um 1150.

78 Vgl. Raible, Wolfgang: Nominale Spezifikatoren (»Artikel«) in der Tradition lateinischer Juristen oder Vom Nutzen der ganzheitlichen Textbetrachtung für die Sprachgeschichte. In: Romanistisches Jahrbuch 36 (1985). S. 44-67. S. 48 f.

79 Stein, Peter: Schriftkultur. Eine Geschichte des Schreibens und Lesens. Darmstadt: WBG, 2006. S. 159 und S. 164-168.

80 Stein (2006), S. 161. S. a. Ludwig, Otto: Geschichte des Schreibens. Bd. 1: Von der Antike bis zum Buchdruck. Berlin: de Gruyter, 2005. S. 112 ff.

81 Vgl. Stein (2006), S. 150 ff.

82 »Ich richte mein Augenmerk auf einen flüchtigen, aber dennoch sehr wichtigen Moment in der Geschichte des Alphabets: den Moment, als – nach Jahrhunderten des christlichen Lesens – die Buchseite sich verwandelte; aus der Partitur für fromme Mürmler wurde der optisch planmäßig gebaute Text für logisch Denkende.« Illich (1991), S. 8.

83 Vgl. Assmann (1994), S. 8.

84 Illich (1991), S. 11.

Scriptio continua

Bis ins frühe Mittelalter hinein waren geschriebene Texte in der so genannten *scriptio continua* verfasst. Die Schrift folgte der grundsätzlich linear organisierten Sprache. Das Aufschreiben von Texten galt lediglich als »graphische Umsetzung vom Oralen ins Literale«,⁸⁵ wobei die Abfolge der Buchstaben in der Zeile der Folge der zu sprechenden Laute entsprach. In diesem Zusammenhang wird ein geschriebener Text gern als eine »Partitur von Lauten« bezeichnet.⁸⁶ Die Buchstaben wurden ohne Leerräume oder andere Zeichen zur Trennung einzelner Wörter aneinandergereiht.

Scriptio continua ist der Versuch, die Erstreckung einer gesprochenen Äußerung in der Zeit mit räumlichen und d. h. sichtbaren Mitteln nachzubilden. Die räumlichen Mittel beschränken sich allerdings auf spezielle Zeichen für die kleinsten lautlichen Einheiten, die Phoneme, sowie auf ihre Anordnung in der Zeile. Deren Richtung zeigt die Erstreckung in der Zeit an.⁸⁷

Der geschriebene Text entsprach damit der mündlichen Tradition und ihren Produktions- und Rezeptionsbedingungen. Der Autor diktierte den Text einem Schreiber, die geschriebenen Texte selber wurden laut oder halblaut gelesen. Dies entsprach dem bis dahin gängigen Bildungskonzept. Das Lesen von Texten war gleichzeitig Meditation und Memoriation. Texte wurden als einheitlicher Rede- bzw. Gedankenfluss verstanden. So wurden sie rezipiert und so wurden neue Texte produziert. Geschriebene Texte galten als sekundäre, der Sprache untergeordnete Erscheinung, eine Einschätzung, die schon bei Platon zu finden ist⁸⁸ und die mit dem Prinzip des phonetischen Alphabets verbunden ist.

Dieses streng lineare Textverständnis war eindimensional. Die aneinandergereihten Laute lassen sich als Linie denken, die dem Verlauf der Zeit folgen. Genauso reihten sich die Buchstaben auf dem Papier aneinander.⁸⁹ Linear, wie der Autor seinen Text als Lautfolge sprach, wurde er vom Schreiber als ununterbrochenes Band von Buchstaben aufgeschrieben. Beim Lesen wurde die kontinuierliche Buchstabenfolge wieder in die kontinuierliche Lautfolge der Sprache umgesetzt.

Ende des 7. Jahrhunderts setzte sich von Irland aus vermehrt die Praxis durch, Sinneinheiten und Wörter durch Leerräume voneinander zu abzugrenzen.⁹⁰ Von der Aufhebung der *scriptio continua* zu sprechen ist eigentlich falsch – Wörter werden (bis heute) zusammenhängend geschrieben. »Die Aufhebung des Prinzips bedeutete zunächst einmal nur, dass nicht mehr unbedingt kontinuierlich geschrieben werden musste, sondern auch Lücken zwischen den einzelnen Zeichen möglich waren.«⁹¹ Die Konsequenzen, die sich aus dieser Möglichkeit ergaben, waren jedoch weitreichend.

2.1.2 Text nimmt materielle Gestalt an

Eine der wichtigsten Eigenschaften von Büchern ist ihre materielle Gestalt. Erst mit dem Aufkommen von Büchern, die dem Text einen festen, materiellen Rahmen verliehen, entwickelten sich die uns geläufigen Formen geschriebener Texte.

Die Strukturierung geschriebener Texte – die zweite Dimension von Texten

Durch die Verwendung von Leerräumen als Trennzeichen zwischen Wörtern kam dem Schriftträger Bedeutung zu. Buchstaben sind phonetische Zeichen. Raum zwischen den Zeichen ist in diesem Sinne kein Zeichen. Die

85 Stein (2006), S. 160.

86 Vgl. Illich (1991), S. 10; Ludwig (2005), S. 60; Stein (2006), S. 160.

87 Ludwig (2005), S. 109.

88 So resümiert Phaidros Sokrates: »Du redest von der lebendigen und beseelten Rede des Wissenden, von der die geschriebene mit Recht ein Abbild genannt werden mag.« Platon: Sämtliche Werke. Berlin: Lambert Schneider, [1940]. Bd. 2. S. 476.

89 Vgl. Raible, Wolfgang: Die Semiotik der Textgestalt. Erscheinungsformen und Folgen eines kulturellen Evolutionsprozesses. Heidelberg: Winter, 1991. Hier besonders S. 5 f.

90 Vgl. Stein (2006), S. 160 ff.

91 Ludwig (2005), S. 112.

Anordnung und Verteilung der Schriftzeichen auf der Fläche eröffnete Möglichkeiten, die nur auf einem zweidimensionalen Schriftträger zu realisieren waren. Es bildete sich eine komplett neue Bedeutungsebene, die bei gesprochenen Texten nicht existierte. Diese neue Bedeutungsebene schriftlicher Texte hatte weder eine lautliche Grundlage, noch konnte sie lautlich realisiert werden.⁹²

Leerräume zwischen den einzelnen Wörtern erleichterten das Lesen. Sie erlaubten das Erkennen von ganzen Wörtern auf »einen Blick« und ermöglichten so neue Lesetechniken, etwa das stille Lesen oder das »Querlesen« und damit das gezielte Auffinden einzelner Wörter oder Textpassagen.

Abstände zwischen Absätzen erlaubten eine weitere Gliederung des geschriebenen Textes. Die Struktur eines Textes wurde deutlich. Auf einzelne Teile des Textes konnte selektiv zugegriffen werden. Wurde ein geschriebener Text vorher als etwas Einheitliches betrachtet (der Text begann mit dem einführenden Incipit, das den folgenden Text vorstellte und gegebenenfalls den Anschluss an andere Texte herstellte und an das sich nahtlos der Text anschloss, der mit dem Explicit bzw. Kolophon ausklang), wurde er nun durch Leerräume in Sinnabschnitte aufgeteilt und damit gegliedert. Einzelne, funktionale Textteile bildeten sich heraus. Neben Absätzen entwickelten sich Überschriften und Glossen. Kurze inhaltliche Angaben, Zusammenfassungen, Argumente oder behandelte Themen eines Abschnittes wurden zunächst am Rand des geschriebenen Textes vermerkt. Durch diese Praxis wurde der Haupttext selber nicht unterbrochen, er wurde noch als etwas Kontinuierliches verstanden und lediglich durch die Anmerkungen am Rand ergänzt. Mit der zunehmenden Segmentierung geschriebener Texte rückten einige dieser Angaben zwischen die Absätze und wurden zu Überschriften.

Die zunächst am Rand untergebrachten Anmerkungen wurden im Laufe der Zeit am unteren Ende einer Seite oder am Ende eines Kapitels oder Buches zusammengefasst und entwickelten sich so zu den heute noch üblichen Apparatformen.⁹³

Der geschriebene Text eignete sich die zweite Dimension an, die die Fläche des Schriftträgers bot. Informationen, die auf ein Trägermaterial geschrieben wurden, konnten sowohl horizontal als auch vertikal gegliedert werden. Die räumliche Verteilung des Textes auf der Fläche des Schriftträgers bildete die Organisations- und Argumentationsstruktur des Textes optisch ab. Es handelte sich damit um eine neue Bedeutungsdimension von Texten, die bei gesprochenen Texten nicht realisiert werden konnte.

Kodizes – die dritte Dimension

Neben der zweidimensionalen Struktur geschriebener Texte gewann die Anordnung der einzelnen Seiten in ihrer Abfolge im Buch Bedeutung. Dem Verständnis von Text als einem linearen Band von Lauten bzw. Gedanken entsprach die Form der Schriftrolle, deren Verwendung bis zum 11. Jahrhundert noch üblich war.⁹⁴ Zwischen Anfang und Ende der Schriftrolle herrschte Kontinuität.⁹⁵ Da nicht von einer Seite auf die nächste geblättert werden musste, gab es keine Unterbrechung des Schriftträgers und des auf ihn geschriebenen Textes. Die Schriftrolle wurde im Zeitraum zwischen dem 4. und 6. Jahrhundert durch den Kodex abgelöst. Im Kodex wurde das Schriftträgermaterial in Bögen aufgeteilt und in Lagen zusammengeheftet. Damit wurde die Kontinuität des Beschreibstoffes unterbrochen und der geschriebene Text auf einzelne Seiten verteilt. Die Seitenstruktur des Kodex kam jedoch erst mit den oben beschriebenen Veränderungen des Textverständnisses zum Tragen. Erst, als man den geschriebenen Text nicht mehr als etwas verstand, das linear vom Anfang bis zum Ende gelesen werden musste, entstand die Möglichkeit, gezielt einzelne Seiten aufzuschlagen und die Ordnungsstruktur

92 Vgl. Ludwig (2005), S. 112 f.

93 Aktuelle »Moden« im Buchsatz greifen zurzeit gern wieder auf die Verwendung von Marginalien zurück, häufig im Rahmen von Layoutkonzepten, die sich an Hypertextfunktionalitäten elektronischer Medien anlehnen.

94 Vgl. Stein (2006), S. 132.

95 Vgl. Raible (1991), S. 6.

der aufeinanderfolgenden Seiten zur Orientierung im Text zu nutzen. Inhaltsverzeichnisse, Register und lebende Kolumnentitel kamen etwa parallel zu den oben beschriebenen Entwicklungen auf.

Die Form des Kodex ergänzte die Schrift um eine dritte Dimension: die aufeinanderliegenden Seiten des Kodex. Sie bildeten zu den zwei Dimensionen der beschriebenen Seite die Tiefendimension. Diese war erst sinnvoll zu nutzen, nachdem die Möglichkeiten der Fläche des Papierträgers entdeckt waren: Inhaltsverzeichnisse waren erst in Verbindung mit Überschriften sinnvoll; Register erst mit der Trennung der einzelnen Wörter durch Leerzeichen, so dass man das entsprechende Wort schnell auf der Seite finden konnte.

Im 13. Jahrhundert hatte das Buch und damit der in ihm tradierte Text seine heute noch gebräuchliche Form gefunden. Auch das Textverständnis hatte sich verändert: Sprache und Text wurden nicht mehr als lineare Einheit betrachtet, sondern als Teile, die einer übergeordneten Struktur folgen.

Dazu kam, dass in dieser Zeit das Papier in Europa Verbreitung fand. Dadurch änderte sich die materielle Qualität der Bücher: Aus schweren, sperrigen Kodizes wurden leichter handhabbare Bücher.

2.1.3 Text und Gestaltung

Neben den Möglichkeiten, die Organisation bzw. die Struktur schriftlicher Texte durch räumliche Aufteilung darzustellen, bildet die Gestaltung der Schrift und des Schriftsatzes einen weiteren wichtigen Aspekt von Texten in Büchern. Die Frage, inwieweit gestalterische Aspekte in Bezug auf die Bedeutung von Texten berücksichtigt werden müssen, wird von der Literaturwissenschaft kaum thematisiert. Erst langsam setzt sich die Erkenntnis durch, dass die »Typographie eines Textes einen Teil von dessen Semantik bilden kann.«⁹⁶

Bei der Gestaltung von Texten werden meist drei Bereiche unterschieden: Typografie, Layout und Design. Die einzelnen Termini werden unterschiedlich, teilweise synonym verwendet. Der Begriff »Typografie« ist ursprünglich eng dem durch Gutenberg eingeführten Buchdruck mit beweglichen Lettern (Typen) verbunden. Im eigentlichen Sinne bestimmt die Typografie das Aussehen der einzelnen Buchstaben (Schriftart, -schnitt, -größe, -farbe) und ihrer Anordnung (Spalten, Zeilenlänge, -abstände, -ausrichtung). Waren dies ursprünglich die wesentlichen Gestaltungsmerkmale von Büchern, wird der Begriff heute auch für sämtliche Aspekte der Buchgestaltung verwendet. Die einzelnen Typen wurden zu Wörtern und Sätzen zusammengesetzt. Die Begriffe »Satz«, »Buchsatz« oder »Drucksatz« werden heute hauptsächlich benutzt, um den Arbeitsschritt bei der Buchherstellung zu beschreiben, in dem das spätere Erscheinungsbild festgelegt wird. Wenn über die Anordnung von Texten, Bildern und grafischen Elementen auf den einzelnen Seiten eines Buches gesprochen wird, benutzt man meist den Begriff »Layout«. Der Begriff Layout stand ursprünglich mit der Technik des Offsetdrucks in Verbindung. Die hierfür als Druckvorlage notwendigen Filme wurden auf großen Leuchttischen »ausgelegt« und zu Druckbögen montiert.⁹⁷ Unter dem Begriff »Design« wird schließlich oft die Gesamtwirkung aus Typografie und Layout zusammengefasst. Er dient auch als Passepartoutbegriff für alle Aspekte der Gestaltung eines Schriftstücks.

Beim Begriff Typografie unterscheidet man den Bereich der Schrift – die Mikrostruktur der Gestaltung, etwa Schriftarten, -größen, -schnitte sowie Buchstaben- Wort- und Zeilenabstände und Sonderzeichen und Ligaturen etc. – von der Makrostruktur des Schriftbildes wie Zeilenbreite, Anzahl der Textspalten, Textausrichtung sowie Einzüge und Abstände von Absätzen. Layout bezieht sich auf die Gestaltung der Buchsseite einschließlich der Anordnung von Bildern, Kolumnentiteln, Paginierung etc.

Die Gestaltung wirkt sich auf verschiedene Ebenen von Texten aus:⁹⁸

96 Nutt-Kofoth, Rüdiger: Text lesen – Text sehen: Edition und Typographie. In: DVjs 78 (2004).S. 3-19. S. 3.

97 Heute wird die Druckvorlage meist am Computer erstellt. Der Film zur Belichtung der Druckplatte wird direkt aus dem Computer heraus belichtet oder entfällt bei modernen CtP-Druckverfahren (Computer to Plate-Druckverfahren) völlig.

98 Vgl. Gumpert, Johann Peter: Zur »Typographie« der geschriebenen Seite. In: Pragmatische Schriftlichkeit im Mittelalter. Erscheinungsformen und Entwicklungsstufen. Hg. von Hagen Keller u. a. München: Fink, 1992. S. 283-292.

- Lesbarkeit – die Gestaltung soll die optimale Lesbarkeit des Textes gewährleisten.
- Textgliederung – gestalterische Merkmale werden zur optischen Gliederung von Texten genutzt.
- Ästhetik – von geschriebenen Texten wird auch optische Gefälligkeit erwartet. Geschriebene Texte haben eine bestimmte Anmutung, die durch die Ästhetik des Textbildes vermittelt wird.
- Semiotik – semiotische Aspekte lassen sich in unterschiedlichen Zusammenhängen darstellen. Zum einen sind Buchstaben selber Zeichen, die neben ihrer phonetischen Bedeutung auch andere Bedeutungsdimensionen haben können, etwa durch ihre Gestaltung. Daneben kann der Anordnung der Buchstaben bzw. der Schrift Bedeutung zukommen. Zu nennen sind hier etwa prosodische oder grafische Aspekte.

Die genannten Ebenen lassen sich nicht klar voneinander abgrenzen. So kann die Lesbarkeit eines Textes durch seine Gliederung beeinflusst werden. Eine sinnvolle Gliederung kann die Orientierung im Text und damit den Zusammenhang und die Argumentationsstruktur verdeutlichen; eine übertriebene Gliederung kann einen Text »zerreißen« und die Lesbarkeit erschweren oder den ästhetischen Eindruck eines Textes stören. Ein ästhetisch angenehmes Schriftbild kann dagegen die Lesbarkeit fördern; ein Übermaß an Zierrat, Schmuckschriften und sonstiger Ausgestaltung kann beim Lesen wiederum störend wirken.

Welche Wirkung die Gestaltung von Text auf der Ebene der Semiotik haben kann, lässt sich bei den dabei möglichen Zusammenhängen und Wechselwirkungen kaum erschöpfend darstellen. Beispielsweise kann sich eine Bedeutungsebene aus der Textästhetik im Zusammenhang mit dem Inhalt ergeben, etwa, wenn ein moderner Text in einer modernen Schriftart gesetzt wird oder bewusst in einer altertümlich anmutenden. Auch das grafische Erscheinungsbild kann eine bedeutungstragende Rolle spielen.

Ein grundlegendes Problem stellt die Abgrenzung zwischen Aufgaben, die den Text selbst betreffen und außertextlichen Bedeutungsebenen dar. Die Frage, was den Text betrifft und was außertextlich ist, hängt wiederum von dem zugrunde gelegten Textbegriff ab. Auf die Problematik, einen Textbegriff zu definieren, habe ich bereits hingewiesen. Auch im Fall der Typografie scheint mir deshalb die Orientierung an der historischen Entwicklung von Schrift und Büchern eine zuverlässige Heuristik zu sein. Auch wenn der Begriff Typografie im engeren Sinne das Gestalten mit reproduzierbarer Schrift bedeutet, verwende ich ihn hier auch für spätmittelalterliche Handschriften.⁹⁹

Entwicklung typografischer Gliederungsmerkmale

Für die Entwicklung der Typografie spielt die Aufhebung der *scriptio continua* eine entscheidende Rolle. Die kontinuierliche Schreibweise erlaubte nur eine Schriftart. Seit dem Ende des 8. Jahrhunderts hatte sich die Karolingische Minuskel nahezu im gesamten westeuropäischen Raum durchgesetzt und war quasi Standard in den Klosterskriptorien und Hofschulen.¹⁰⁰ Vor dem Hintergrund der mittelalterlichen Praxis, Bücher durch Schreiber, denen der Text diktiert oder vorgelesen wurde, niederschreiben oder kopieren zu lassen, stand der Aspekt der Einheitlichkeit und der Lesbarkeit schriftlicher Texte im Vordergrund. Ein gleichmäßiges, ruhiges Schriftbild entsprach am ehesten der meditativ-murmelnden bzw. laut-lesenden Rezeptionsweise sowie der Betrachtung eines Textes als Einheit.

Die Veränderungen im Verhältnis zu geschriebenen Texten im 12. Jahrhundert hatten deutliche Auswirkungen auf ihre schriftliche Gestaltung. Mit der Auftrennung des kontinuierlichen Schriftflusses ging, neben den Möglichkeiten der räumlichen Anordnung der Buchstaben, auch die Möglichkeit einher, verschiedene Schriftarten zu mischen. So ergänzte man zunächst die Zeichen der Karolingischen Minuskel durch die antiken

⁹⁹ Vgl. Gumbert (1992).

¹⁰⁰ Vgl. Stein (2006), S. 152. Stein bezeichnet die Karolingische Minuskel als »Hauptschrift des Mittelalters«.

Vorbildern entlehnte *capitalis rustica*. Dieser kam vornehmlich der Status einer Auszeichnungsschrift zu.¹⁰¹ Sie wurde zum einen für wichtige Wörter oder Abschnitte verwendet. Ein in *Capitalis* geschriebener Buchstabe kennzeichnete etwa den Anfang eines Kapitels oder wichtige Wörter in einem Satz. Ebenfalls vor dem Hintergrund der Aufhebung der *scriptio continua* entwickelten sich Zeichen, die vorher teilweise als »Lesemarkierungen« verwendet worden waren, zu Satzzeichen. Der Beginn eines neuen Satzes wurde ebenfalls durch einen Buchstaben in *Capitalis* gekennzeichnet. Es entstand unsere heutige Groß-/Kleinschreibung.

Zur Darstellung der Textgliederung wurden neben der räumlichen Aufteilung des Textes auf dem Papier und der Verwendung von Auszeichnungsschriften noch weitere Gestaltungsmittel verwendet. Für Überschriften wurden größere Schriftgrößen und andere Schriftfarben verwendet; Anmerkungen und Kommentare wurden dagegen in einer kleineren Schrift gehalten. Petrus Lombardus bediente sich Mitte des 12. Jahrhunderts bereits eines umfangreichen typografischen Instrumentariums. Der Platz für Glossen wurde bemessen, Zitate als solche gekennzeichnet und mit Quellenangaben versehen, Schlüsselwörter wurden teilweise rot unterstrichen.¹⁰²

»Ab der Mitte des 12. Jahrhunderts wird die innere Organisation von Texten [...] durchgehend sichtbar gemacht, ab dem 13. Jahrhundert spiegelt sich die innere Organisation des Textes, die so genannte *ordinatio*, in der Textgestalt wieder.«¹⁰³ Hatte man bisher beispielsweise das Incipit und Explizit durch größere Schrift hervorgehoben und den Anfang einer neuen Geschichte durch eine Initiale gekennzeichnet, fand die zunehmende Fragmentierung und Gliederung der geschriebenen Texte in der Gestaltung Ausdruck.

Eine weitere Schriftart wurde mit der Verbreitung von Papier in Europa wiederentdeckt: Die römische Kursive. Papier war im Vergleich zu Pergament ein schnelles Schreibmaterial. Die Tinten, mit denen auf Papier geschrieben wurde, trockneten schnell und ermöglichten dem selbstschreibenden Autor ein erheblich schnelleres Notieren seiner Gedanken als das Pergament.¹⁰⁴

Typografische Gliederungsmerkmale etablieren sich im Zusammenhang mit der Entstehung und Differenzierung weiterer (geschriebener) Textgattungen. Die typografische Gliederung ermöglicht in den meisten Fällen auf einen Blick die Bestimmung der Textsorte: »Denn für den Leser findet die erste Wahrnehmung des Textes – vor Beginn des Lesens – immer durch das Sehen statt, d. h. im Wahrnehmen der typografischen Struktur des Textes.«¹⁰⁵

Urteile über die Zugehörigkeit eines Textes zu einer Textsorte werden nicht nur aufgrund inhaltlicher oder sprachlich-stilistischer Textmerkmale gefällt. Die Vorstellungen über Textklassen und ihre charakteristischen Eigenschaften umfassen immer auch Merkmale ihrer äußeren (typographischen) Form; sie gehören zum Allgemeinwissen unserer Lesekultur.¹⁰⁶

Aufgrund typografischer Gliederungsmerkmale, die auch als »Dispositive«¹⁰⁷ bezeichnet werden, kann ein Leser erkennen, ob es sich bei dem vor ihm liegenden schriftlichen Text um ein Gedicht, ein Drama oder einen Lexikonartikel handelt. Dispositive lassen sich einerseits zum Aspekt der Textgliederung zählen, weil die meisten Textsorten eine spezifische Gliederungsstruktur besitzen; zum anderen haben sie auch eine semantische Dimension, nämlich die, dass man ohne den Text selber gelesen zu haben bereits anhand des Dispositivs eine Information über die Textgattung erhält. Die Grundlagen zur Entwicklung solcher Dispositive lassen sich bis ins 12. Jahrhundert zurückverfolgen.

101 Vgl. Ludwig (2005), S. 116 f.

102 Vgl. Illich (1991), S. 105.

103 Raible (1991), S. 10.

104 Vgl. Ludwig (2005), S. 200 ff.

105 Nutt-Kofoth (2004), S. 3.

106 Wehde (2000), S. 119.

107 Vgl. Wehde (2000), S. 119 sowie Chartier, Roger: Lesewelten. Buch und Lektüre in der frühen Neuzeit. Frankfurt a. M./New York: Campus, 1990. S. 8.

Semantische Aspekte von Typografie

Semantische Aspekte werden in ihrer Bedeutung für die Rezeption schriftlicher Texte selten thematisiert. Sie lassen sich schwer beschreiben und ihre Wirkung hängt stark vom Leser und dessen Wissenshintergrund ab.

Die zur Gliederung und Strukturierung von Texten verwendeten Aspekte stellen auf der semantischen Ebene Selektionskriterien dar. Sie schaffen Bedeutung, indem sie zeigen, welche Teile des Textes zusammengehören und Sinneinheiten bilden. Ebenso zählen die erwähnten Dispositive zu den semantisch relevanten Aspekten von Typografie. Sie können einem Leser doch Informationen über einen Text vermitteln, indem sie beispielsweise in vielen Fällen eine Gattungszuordnung ermöglichen (s. o.).

Die verwendete Schriftart kann ebenfalls eine semiotische Dimension haben. In Büchern, die in Fraktur gesetzt sind, wird oft eine Antiqua-Schrift verwendet, um Fremdwörter oder fremdsprachige Textteile zu kennzeichnen.

Schließlich kann die Form, in der der Text gesetzt ist, eine Bedeutung haben. Als Beispiele hierfür lassen sich Figurengedichte, konkrete Poesie oder unterschiedliche Verwendungsarten von Schrift in der Werbung anführen.

Hier alle möglichen Aspekte und Zusammenhänge aufzuzeigen, ist schier unmöglich, zudem sind viele Aspekte subjektiv oder bisher selten untersucht und beschrieben worden. Der Bereich der Werbung und der visuellen Kommunikation macht sich solche Phänomene zu Nutze. Es gibt viele Empfehlungen, wie eine Typografie einzusetzen ist. Ein zuverlässiges wissenschaftliches Instrumentarium, um konkrete Phänomene zu beschreiben, fehlt bislang.¹⁰⁸

Letztendlich lässt sich anscheinend nur recht allgemein konstatieren, »dass die Formen den Sinn erzeugen und ein schriftlich verankerter Text eine neue Bedeutung und einen neuen Status erhält, wenn sich die Dispositive des typografischen, zum Lesen bestimmten Objekts verändern«.¹⁰⁹

Die folgenden vier Beispiele sollen lediglich einen Eindruck davon vermitteln, wie und auf welchen Ebenen Aspekte von Information und Gestaltung bei Texten miteinander verbunden sein können.

Die Bedeutung des Wortes »Reich« in Abhängigkeit von der Schriftart

Das Wort »Reich« kann, unterschiedliche Bedeutungen haben, beispielsweise:

- Land oder Länder, das bzw. die von einem Monarchen regiert werden
- bis zum Jahre 1945 Synonym für Deutschland
- Bereich der materiellen oder geistigen Welt
- Biologie: hohe taxonomische Kategorie (Regnum)
- ...

Das Wort »Reich«, in unterschiedlichen Schriftarten gesetzt, kann bei Lesern unterschiedliche Assoziationen hervorrufen:

Reich in einer gebrochenen Schrift gesetzt, hier in einer Schwabacher, lässt schnell Gedanken an ein deutsches Reich aufkommen, sei es im Sinne des Nationalstaatsgedankens des 19. Jahrhunderts oder (fälschlicherweise) der nationalsozialistischen Ideologie.¹¹⁰

Reich in einer Jugendstil-Schriftart, hier der Arnold Böcklin, kann dagegen Gedanken an ein künstlerisches Reich, ein dichterisches Arkadien aufkommen lassen.

¹⁰⁸ Wehde (2000) unternimmt zwar den Versuch, genau solche Phänomene zu beschreiben, gelangt aber nicht über eine Aufstellung der zur Verfügung stehenden Theorien und der Beschreibung typografischer Phänomene hinaus.

¹⁰⁹ Chartier (1990), S. 8.

¹¹⁰ Nachdem die Nationalsozialisten Frakturschriften zunächst als »deutsche Schrift« propagiert hatten, wurden sie am 3. Januar 1941 durch den so genannten »Normalschrifterlass« verboten.

Wie das Wort »Reich« in Verbindung mit unterschiedlichen Schriftarten interpretiert wird, hängt stark vom Bildungshintergrund des Lesers ab. Wer eine Arnold Böcklin-Schrifttype nicht als Jugendstil-Schriftart erkennt und mit der Literatur des Jugendstils nicht vertraut ist, kann auch nicht zu der angeführten Assoziation des Wortes »Reich« kommen. Was also das Ergebnis solcher Assoziationen ist, lässt sich nicht sicher voraussagen. Dass ein Leser solche Assoziationen aber unwillkürlich hat, wenn er Text liest, lässt sich kaum bezweifeln.

Es scheint problematisch, solche Phänomene konkret zu erfassen und zu beschreiben, da sie zum einen zu viele subjektive Aspekte berühren, wie etwa das »Weltwissen« des Lesers, zum anderen sind sie auch einem modischen Wandel unterworfen. War die Ästhetik der 1960er Jahre in ihrer Zeit modern und drückten sich in ihr Visionen der Zeit aus, so war sie in den 1980er Jahren vollkommen »out«; zwanzig Jahre später dagegen als »Retro« wieder modern. Ähnliche eklektizistische Tendenzen lassen sich in verschiedenen Epochen des Buchdrucks finden.

Text im Text in Lessings frühem Drama DIE ALTE JUNGFER

In Lessings Drama DIE ALTE JUNGFER findet sich im dritten Aufzug, dritter Auftritt, eine Szene, in der die Figur des Herrn Kräusel ein eigenes Gedicht vorliest und kommentiert. Die Figur des Herrn Kräusel ist als Parodie auf populäre Dichter der Zeit angelegt.

Da es sich bei dem Text um ein Drama handelt, ist er darauf ausgelegt, gesprochen zu werden. Der Text selber soll gesprochene bzw. zu sprechende Sprache sein. In den zu sprechenden Text ist ein Textbereich eingefügt, der gesprochen werden soll, als ob er gelesen wird. Innerhalb des auf Mündlichkeit ausgelegten Dramas sollen Schriftlichkeitsmerkmale zum Ausdruck gebracht werden, damit ein Schauspieler weiß, dass er diesen Text sprechen soll, als ob er gelesen wird bzw. einem Leser des Dramas auffällt, dass hier ein Text im Text vorliegt. Man könnte sagen, dass die schriftliche Darstellung darauf abzielt Mündlichkeitsmerkmale darzustellen, die in der mündlichen Sprache wieder Schriftlichkeitsmerkmale transportieren sollen.

Nahezu alle verfügbaren Ausgaben dieses Dramas versuchen dieser Textstruktur, in der sich Mündlichkeitsmerkmale und Schriftlichkeitsmerkmale mischen, durch eine besondere typografische Darstellung gerecht zu werden.

In der Erstausgabe des Dramas¹¹¹ fällt der Auftritt dadurch auf, dass er nicht dem üblichen Schema der Absatzzeile folgt. Die Zeilen des Gedichts sind linksbündig gesetzt, die Kommentare Kräusels dagegen eingezogen, wobei der erste mündliche Kommentar (»Das nenne ich schalkhaft...«) aus der Reihe fällt und ebenfalls linksbündig gesetzt ist, vermutlich, um unter der Initiale kein »Loch« entstehen zu lassen. Doch auch die Kommentare Kräusels entsprechen in ihrem Einzug nicht dem in der Ausgabe üblichen Einzug von Sprechertext, vielleicht, weil der Sprecher hier nicht unmittelbar vor den Sprechertext gesetzt, sondern in der darüberstehenden Szenenanweisung genannt ist.

In der ersten Auflage der von Karl Lachmann 1838 herausgegebenen SÄMTLICHEN SCHRIFTEN LESSINGS¹¹² werden die Einzüge umgekehrt gehandhabt: Hier ist das Gedicht eingerückt, Kräusels Kommentar steht dagegen linksbündig. Die »gesprochenen« Passagen folgen damit der typografischen Anordnung gesprochener Passagen in den übrigen Teilen des Dramas. Das Gedicht ist zusätzlich in einer fetteren Schriftart gesetzt und somit im Schriftbild hervorgehoben. Der verwendete Schriftschnitt findet in dieser Größe sonst bei den Szenenanweisungen Verwendung.

111 Lessing, Gotthold Ephraim: Die alte Jungfer. Ein Lustspiel in drey Aufzügen. Berlin: 1749. Wiedergabe in: Bibliothek der deutschen Literatur. Eine Ed. der Kulturstiftung der Länder. Mikrofiche-Gesamtausg. nach den Angaben des Taschengoedeke. München u. a.: Saur, [1994]. 19867.

112 Lessing, Gotthold Ephraim: Gotthold Ephraim Lessings sämtliche Schriften. Hg. von Karl Lachmann. Berlin, 1838 ff. Bd. 3. S. 408 f.

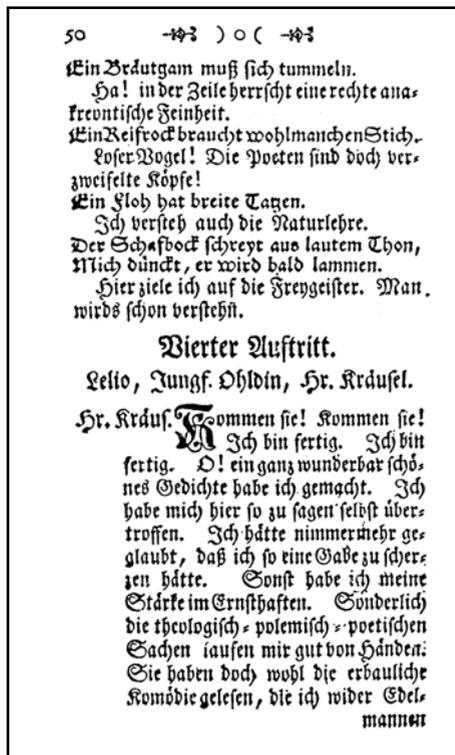
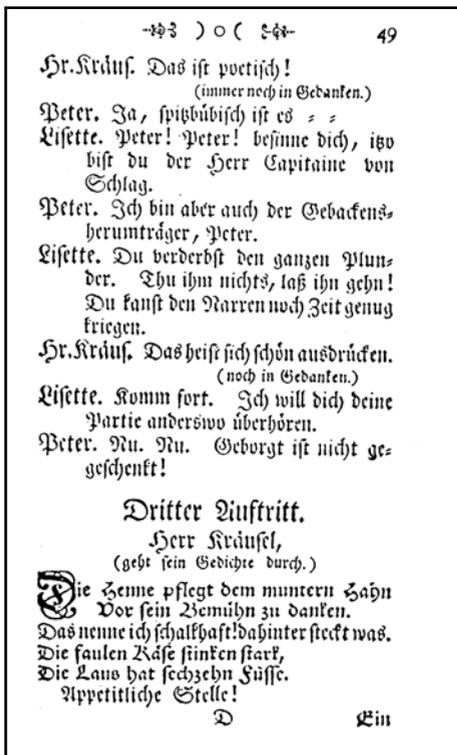


Abbildung 11. Lessing: Die alte Jungfer in der Erstausgabe von 1749

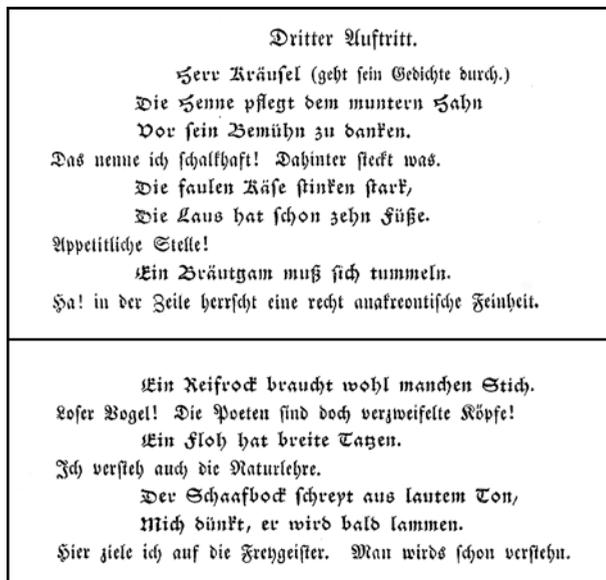


Abbildung 12. Lessing: Die alte Jungfer in den Sämtlichen Schriften, 1. Auflage von 1838

In der dritten, durchgesehenen Auflage von 1887¹¹³ wird die Stelle vollkommen anders wiedergegeben. Hier steht das Gedicht linksbündig, der Kommentar ist eingezogen, zusätzlich ist der Gedichtstext gesperrt gedruckt. Der gesperrte Druck der Passagen, in denen Herr Kräusel sein eigenes Gedicht liest, kann als prosodisches Merkmal betrachtet werden, als Versuch, typografisch darzustellen, wie Herr Kräusel den eigenen Text Buchstabe für Buchstabe liest. Die typografische Gestaltung kann als Hinweis verstanden werden, dass ein Schauspieler den Text quasi buchstabierend zu sprechen hat, was die Figur des Herrn Kräusel auf eine bestimmte Weise näher charakterisieren würde: als »Poet«, der selber Schwierigkeiten beim Lesen hat, was darauf schließen ließe, dass ihm auch das Schreiben nicht leicht fällt, was sich wiederum auf einer anderen Ebene durch die Qualität seines Gedichts bestätigt. Oder es lässt auf die Eitelkeit des Herrn Kräusel schließen, der beim Lesen seines Gedichts nochmals jedem einzelnen Buchstaben Bedeutung verleihen will, eine Interpretation, die auch durch seine Kommentare zu seinem eigenen Gedicht gestützt wird.

Eine besondere typografische Hervorhebung findet sich in nahezu allen Editionen, so auch in moderneren, in Antiqua gesetzten Ausgaben wie der Werkausgabe des Aufbau Verlags¹¹⁴ oder der dreibändigen Lessing-Ausgabe bei Artemis und Winkler.¹¹⁵ Dabei werden die Einzüge unterschiedlich gehandhabt: In der Aufbau-Ausgabe ist das Gedicht eingezogen und der Kommentar Kräusels linksbündig, bei Artemis und Winkler verhält es sich umgekehrt.



Abbildung 13. Lessing: Die alte Jungfer in den Sämtlichen Schriften, 3. Auflage von 1887

Der linksbündige, gesperrte Satz wie in Abbildung 13 betont dabei eher den mündlichen Aspekt des gelesenen Gedichts: Der linksbündige Satz markiert den Haupttext, die gesperrte Schrift betont den gelesenen Charakter des Sprechtextes. Die Kommentare wirken optisch in das Gedicht eingeschoben. Der umgekehrte Fall, in dem das Gedicht eingezogen ist (Abbildung 12), erweckt den Eindruck, dass das Blatt mit dem Gedicht im Text sichtbar wird. Hier wird der Aspekt der Schriftlichkeit des Gedichts typografisch betont. Einige Ausgaben unterstützen diese Wirkung zusätzlich dadurch, dass sie das Gedicht in einer kleineren Schrift setzen. Ein besonderes Beispiel

113 Lessing, Gotthold Ephraim: Gotthold Ephraim Lessings sämtliche Schriften. Hg. von Karl Lachmann. 3., auf's neue durchges. und verm. Aufl.. Besorgt durch Franz Muncker. Stuttgart: Göschen, 1887. Bd. 3. S. 224.

114 Lessing, Gotthold Ephraim: Gesammelte Werke. Berlin: Aufbau, 1954. Bd. 3. S. 466.

115 Lessing, Gotthold Ephraim: Werke in drei Bänden. 3. Aufl. München: Artemis u. Winkler, 1994. Bd. 1. S. 201.

bildet die Anfang des 20. Jahrhunderts erschienene, von Franz Boxmüller herausgegebene Ausgabe,¹¹⁶ die die Zeilen des »geschrieben« Gedichts zusätzlich zur kleineren Schrift noch in Anführungszeichen setzt, entweder, um es als eine Art Zitat zu kennzeichnen oder um dem Schriftbild des Gedichts wieder Mündlichkeitsmerkmale zu verleihen.

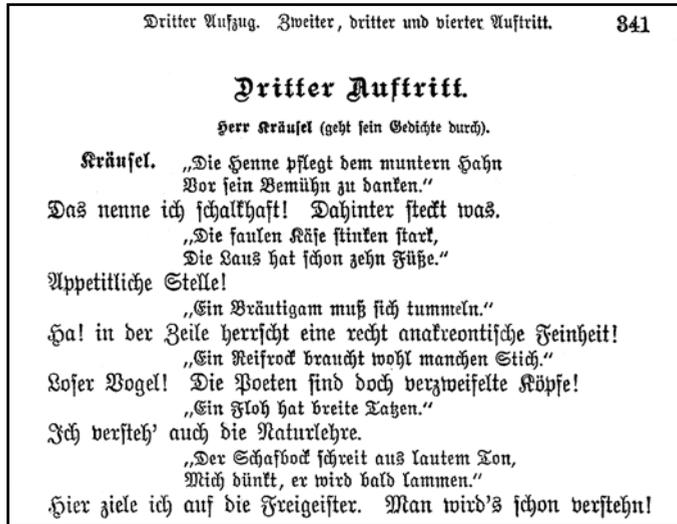


Abbildung 14. Lessing: Die alte Jungfer in Lessings Werke. Hg. von Franz Boxmüller

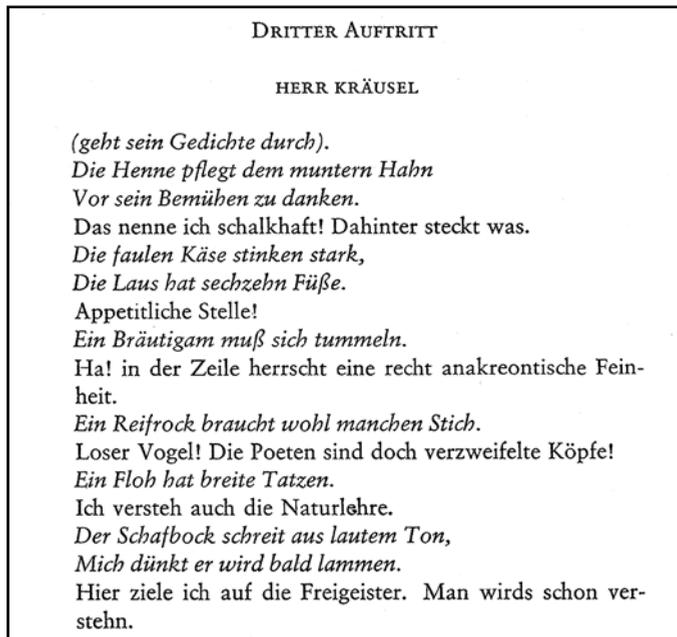


Abbildung 15. Lessing: Die alte Jungfer in der Münchner Hanser-Ausgabe

116 Lessing, Gotthold Ephraim: Lessings Werke. Hg. von Franz Boxmüller. Leipzig u. Wien; Bibliographisches Institut, [1908]. Bd. 1. S. 341.

In nahezu allen Ausgaben, die ich zum Vergleich heranziehen konnte, fällt auf, dass die verwendeten Einzüge generell von den im übrigen Drama verwendeten abweichen. Offensichtlich stellt die Textpassage besondere typografische Anforderungen.

Eine eigentümliche Darstellung des Textes wählte der Hanser-Verlag.¹¹⁷ Hier ist das Gedicht kursiv gesetzt, gesonderte Einzüge gibt es nicht. Kursivierungen werden an anderen Stellen dieser Ausgabe für Regieanweisungen verwendet. Ein möglicher Grund könnte sein, dass man sich an einer Fraktur-Ausgabe orientiert hat, in der das Gedicht gesperrt gesetzt ist. Für die Antiqua-Ausgabe wurde der gesperrte Text der Frakturausgabe kursiv wiedergegeben, was bei der Umsetzung von Frakturschriften in eine Antiqua gängige Praxis ist. Eine andere Erklärung wäre, dass man sich an der Erstausgabe orientiert hat. Wie oben erwähnt, werden in der Erstausgabe für das Gedicht der Schriftschnitt und die Schriftgröße verwendet, die auch für die Regieanweisungen verwendet werden. Insofern wäre es konsequent, wenn dies auch in einer neueren Ausgabe so gehandhabt würde. Das Beispiel zeigt, wie typografische Gestaltungsmöglichkeiten Informationen transportieren können, die über die Informationsebene der reinen Buchstaben- bzw. Wortbedeutung hinausgehen und die schließlich auch zu unterschiedlichen Bewertungen bzw. Interpretationen des Textes führen können.

Ein Spiel mit Bedeutungen und Buchstaben. FAUST. EINE TRAGÖDIE von Johann Wolfgang von Goethe

Ein Beispiel, bei dem das Aussehen von Buchstaben auf einer ganz anderen Ebene eine Rolle spielt, findet sich in Goethes FAUST. DER TRAGÖDIE ZWEITER TEIL. Mephistopheles bezeichnet die »Greife« als »Greise«, wogegen sich diese wehren.

Die Verwechslung der Wörter »Greife« und »Greise« wird auch im Schriftbild plausibel, wenn die Textstelle in einer Frakturschrift gesetzt ist, die ein Mittel-s verwendet: **Greife** und **Greise** klingen in diesem Fall nicht nur sehr ähnlich; sie sehen auch zum Verwechseln ähnlich aus, da sich in einer Frakturschrift die Typen des **f** und des **f** kaum unterscheiden.

Im Gegensatz dazu ist in einer Antiqua-Schrift, die überdies nicht über ein Mittel-s verfügt, eine Verwechslung der Wörter schwieriger. Die Verwechslung der Wörter »Greife« und »Greise« wird nicht erst ersichtlich, wenn man die Wörter ausspricht, also laut liest, sondern ist aufgrund der aufgezeigten typografischen Besonderheit in einer Frakturschrift schon beim stummen Lesen angelegt. Hier wird auf der typografischen Ebene ein Missverständnis dargestellt, das ursprünglich der sprachlichen Ebene angehört. Die Möglichkeit der Darstellung ist in diesem Fall von der Schriftart abhängig. Nur in einer Frakturschrift, die über ein Mittel-s verfügt, ist die Ähnlichkeit der phonetischen Qualität ähnlich hoch wie bei der grafischen Qualität des Buchstabens.

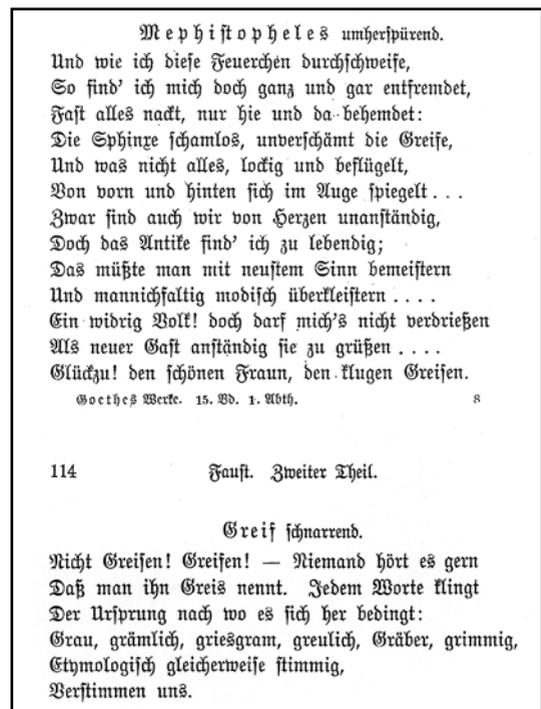


Abbildung 16. Ausschnitt aus Goethes Werke, hg. im Auftrage der Großherzogin Sophie von Sachsen. Weimar: Hermann Böhlau, Erste Abt., Bd. 15, 1888, S. 113 f

117 Lessing, Gotthold Ephraim: Werke. Hg. von Herbert G. Göpfert. Bd. 1-8, München: Hanser, 1970 ff. S. 762.

Bildwirkung des Schriftbildes

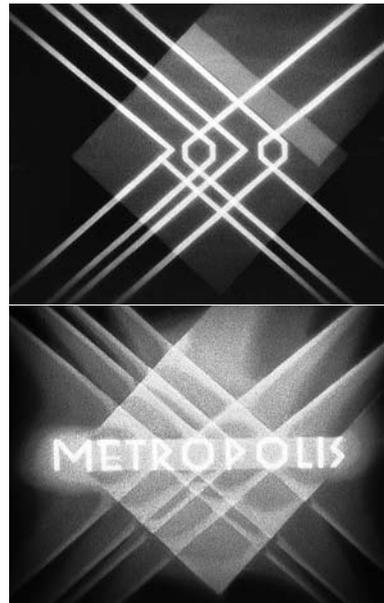
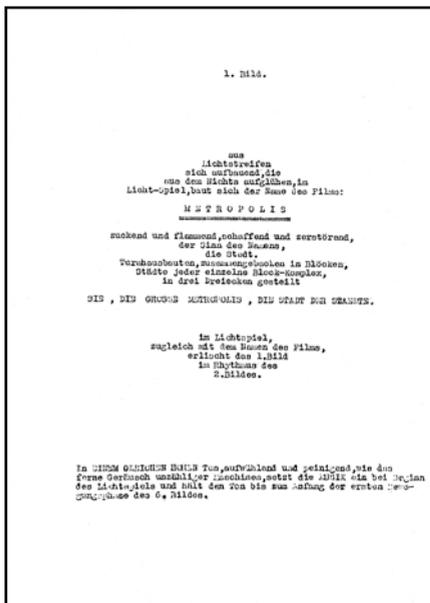
Auf der Ebene der Absatz- bzw. Seitengestaltung können auch grafische Aspekte zum Tragen kommen. Recht bekannt sind auf diesem Gebiet Figurengedichte und Werke der konkreten Poesie, beispielsweise »Die Trichter« von Christian Morgenstern oder Reinhard Döhls »Apfel«.

Die Trichter ¹¹⁸

Zwei Trichter wandeln durch die Nacht.
Durch ihres Rumpfs verengten Schacht
fließt weißes Mondlicht
still und heiter
auf ihren
Waldweg
u. s.
w.

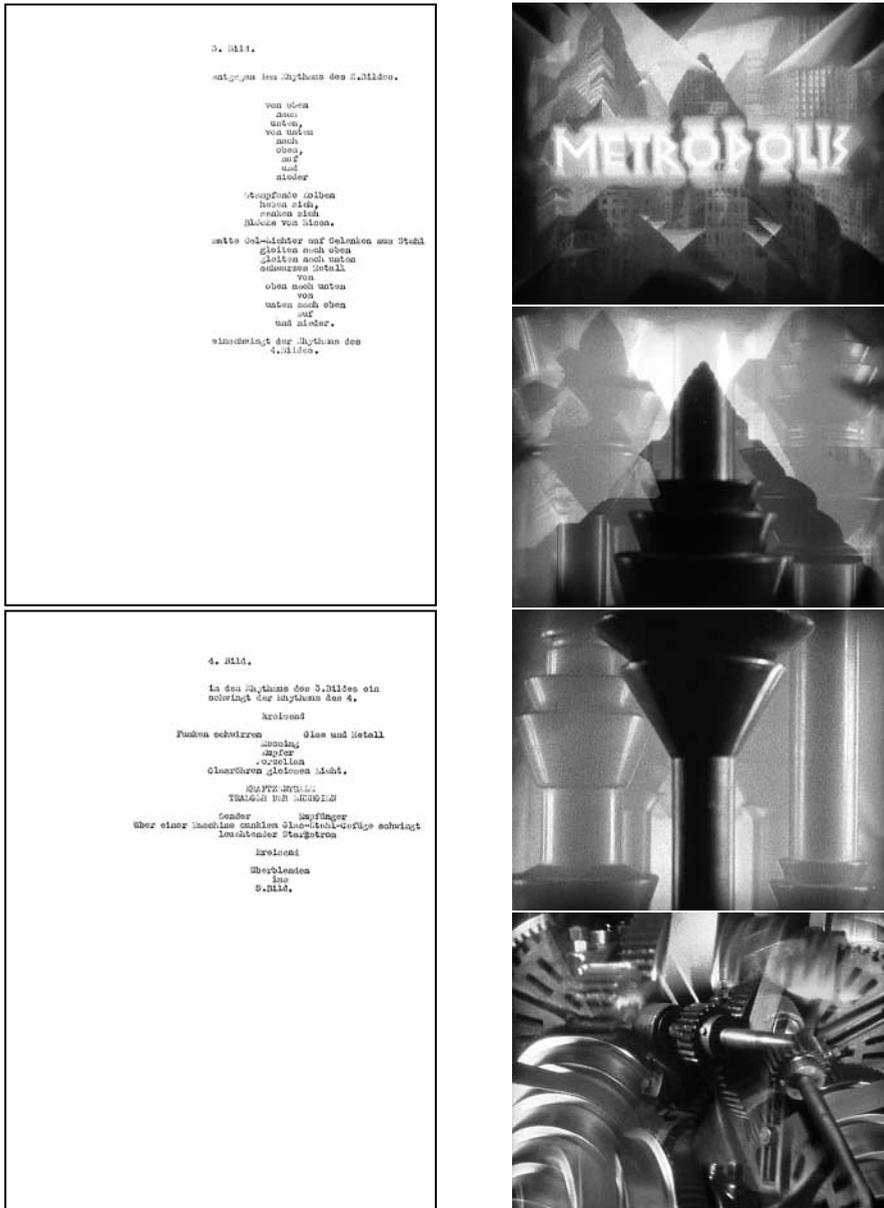
Typografie kann jedoch nicht nur benutzt werden, um einen Gegenstand nachzubilden, sondern auch, um eine Form, ggf. sogar eine Bewegung vorzugeben. Ein Beispiel ist die Eingangssequenz des Stummfilmklassikers METROPOLIS.¹¹⁹ Das Schriftbild im Drehbuch von Thea von Harbou gibt in Form und Abfolge die Formen und Metamorphosen der Filmsequenz vor:

Die Typografie des Drehbuchs nimmt hier die Form des späteren Filmbildes vorweg. Durch die Gestaltung des Textes werden zusätzliche Informationen (die Formensprache der im Text beschriebenen Bildsequenzen) mit Hilfe der Typografie beschrieben bzw. gezeichnet. Hier geht der Informationsgehalt der Typografie über den des Textes hinaus.



118 Morgenstern, Christian: Alle Galgenlieder. Zürich: Diogenes, 1981. S. 30.

119 Metropolis. Deutschland: Ufa, 1927. Regie: Fritz Lang, Drehbuch: Thea von Harbou.

Abbildung 17. Paralleldarstellung von Drehbuchtext und Standbildern aus der Anfangssequenz von Metropolis¹²⁰

120 Das hier abgebildete Drehbuch-Typoskript ist im Besitz der Stiftung Deutsche Kinemathek Berlin. Die screenshots stammen von der DVD Fritz Lang's Metropolis. Neu restaurierte Fassung mit Originalmusik. München: Transit Film, 2003.

Ästhetik der Typografie

Die ästhetische Dimension von Schrift beschränkte sich, wie oben erwähnt, zuerst auf ein einheitliches und ruhiges Schriftbild. Die Ausgestaltung von Büchern an sich ist, besonders im Hinblick auf das Umfeld, für das sie geschrieben wurden, von Bedeutung. Bereits im Mittelalter gab es aufwendig gestaltete Bücher für repräsentative Zwecke in Kirchen und an Fürstenhöfen, die mit künstlerisch ausgestalteten Initialen, bunten Randbemalungen und Goldverzierungen versehen waren. Daneben gab es aber auch »Studienausgaben«, bei denen auf reichhaltige Ausstattung verzichtet wurde und die z. B. für den Studienbetrieb der Bettelorden wie Franziskaner und Dominikaner gedacht waren, die sich damals bewusst gegen den Prunk der Kirche stellten.

Indem das Buch sich teurer, billiger, moderner, konservativer als andere zeigt, bestätigt es seinem Besitzer den Anspruch, mächtiger, tugendsamer, aufgeschlossener, weiser zu sein und hilft ihm, diesen Anspruch anderen gegenüber geltend zu machen.¹²¹

Ein unmittelbarer Zusammenhang zwischen einem Text bzw. seinem Inhalt und seiner ästhetischen Ausgestaltung lässt sich dabei aber schwer nachweisen. Das Schreiben von Texten hat einen eher handwerklichen Charakter, der in besonderen Fällen in Richtung Kunsthandwerk geht.

Mit der Verbreitung des Buchdrucks änderte sich zunächst wenig. Es fällt eher auf, dass gedruckte Bücher sich an handschriftlichen Traditionen orientierten.¹²² Abkürzungen, wie sie zum schnelleren handschriftlichen Schreiben eingeführt wurden, wurden im Druck übernommen, obwohl sie hier ihren ursprünglichen Sinn, das Schreiben zu beschleunigen, nicht mehr erfüllten. Vielmehr bedeutete ihre Verwendung, dass eine größere Vielfalt von Typen hergestellt werden musste,¹²³ was einen größeren Aufwand bedeutete als die Beschränkung auf die normalen Buchstaben. Auch die Seitengestaltung orientierte sich an handschriftlichen Vorbildern. Hier wurde ebenfalls ein hoher Aufwand betrieben, um die von Handschriften gewohnte Anmutung der Texte zu wahren. So wurden früh Versuche unternommen mehrere Farben zu drucken.¹²⁴ Was bei handschriftlich vervielfältigten Texten kein größeres Problem darstellte, da die Verwendung verschiedenfarbiger Tinten, meist Schwarz und Rot, lediglich den Wechsel der Feder verlangte, bedeutete für den Druck einen nicht unerheblichen Aufwand.

Bücher und Schrift erfuhren im weiteren Verlauf der Entwicklung immer wieder künstlerische Ausgestaltungen. Es entstand eine Buchkunst, die jedoch als eigenständiges Phänomen zu betrachten ist, da sie in keiner unmittelbaren Wechselwirkung mit dem Text stand. Dies wird besonders daran deutlich, dass es die gleichen Texte in unterschiedlichsten Ausgestaltungen gibt, man denke etwa an verschiedene Bibelausgaben, ohne dass sich dadurch etwas in Bezug auf den Text selber ändert.

Einen Einschnitt bedeutete in diesem Zusammenhang vielleicht die Buchkunstbewegung. Am Ende des 19. Jahrhunderts von England ausgehend, fand sie besonders in Deutschland vor dem Hintergrund des Jugendstils und der Idee des »Gesamtkunstwerks« Niederschlag. Beispielhaft können hier etwa Stefan George und die Schriftsteller aus dem »George-Kreis« angeführt werden, die die sezessionistische Absicht ihrer Kunst auch typografisch durch generelle Kleinschreibung, Verwendung einer eigens entworfenen latinisierenden Type und kostbarer künstlerischer Ausstattung betonten.

In den meisten Fällen lässt sich aber ein engerer Zusammenhang zwischen der Ästhetik und dem gedruckten Text nicht herstellen. »Ein wirklich gemeinsamer Schaffensprozess zweier gleichwertiger Künste ist nicht ein einziges Mal nachweisbar. Schließlich musste die ›Synthese der Kunst‹ eine Utopie bleiben.«¹²⁵ Die ästhetische Gestaltung eines schriftlichen Textes kann auf der semantischen Ebene lediglich als Indiz dafür herangezogen

121 Gumpert (1992), S. 284.

122 Vgl. Chartier (1990).

123 Für die 42-zeilige Bibel (B42) hat Gutenberg 290 Typen verwendet, um im Druck den Charakter der Handschrift zu bewahren. Vgl. Funke, Fritz: *Buchkunde. Die historische Entwicklung des Buches von der Keilschrift bis zur Gegenwart*. 6. überarb. und erg. Aufl. Wiesbaden: Albus, 2006. S. 105.

124 So wurde im Mainzer Psalter von 1457 versucht, die Initialen zweifarbig – rot und blau – zu drucken. Vgl. Funke (2006), S. 107.

125 Pantus, Willem-Jan: *Jugendstil in Wort und Bild. Illustrierte Dichtkunst um 1900*. Köln: LETTER Stiftung, 2000. S. 173.

werden, wann der Text gedruckt wurde. Da aber genauso z. B. klassische Texte in Jugendstilausgaben herausgegeben wurden, lassen sich nicht einmal Rückschlüsse auf den Entstehungszeitraum des Textes selber ziehen.

Wie problematisch es ist, einen Text aufgrund der typografischen Ästhetik einzuordnen, zeigt schließlich die Verwendung von Frakturschriften. Vielfach wird heute ein in einer Frakturschrift gesetzter Text mit propagiertem Deutschtum, insbesondere mit dem Nationalsozialismus, in Verbindung gebracht. Dabei war es gerade die Reichsleitung des Naziregimes, die Frakturschriften 1941 als »Schwabacher Judenlettern« verunglimpfte und die Umstellung auf »Normal-Schrift«, d. h. Antiqua-Schriften, verordnete.

2.1.4 Buchproduktion und Distribution

Der Buchdruck hatte zunächst nur geringe Auswirkungen auf Form und Gestaltung schriftlicher Texte. Vielmehr orientierte man sich an den Vorbildern handgeschriebener Bücher bzw. Kodizes. Die Bedeutung des Buchdrucks liegt vielmehr in den Möglichkeiten der mechanischen Vervielfältigung von Texten und damit ihrer massenhaften Verbreitung.

Die Einführung des Buchdrucks ist der Beginn eines Prozesses der Verfügbarmachung und Verbreitung schriftlicher Texte. Die technische Entwicklung im Bereich der Drucktechnik erlaubte die Herstellung von immer mehr Druckerzeugnissen in immer kürzerer Zeit. Zu diesen Entwicklungen gehörten neben der Weiterentwicklung der Druckmaschinen auch Techniken des Buchsatzes (Guss von Lettern, Satzmaschinen), des Falzens und Bindens von bedrucktem Papier zur Buchform und nicht zuletzt Verbesserungen bei der Produktion von Papier.

Damit einhergehend verändern sich Distribution und Rezeption der Druckwaren.

Spätestens Ende des 19. Jh. waren Buch- und Pressewesen produktionstechnisch, distributiv und organisatorisch so aufgestellt, dass sie ein Millionenpublikum bei weiter wachsender Bevölkerungszahl problemlos mit allen möglichen Sorten von Lesestoff versorgen konnten.¹²⁶

Damit waren die Grundlagen unseres heutigen Literaturbetriebes geschaffen. Die technische Entwicklung ist indes nicht stehen geblieben. Auf dem Gebiet der Drucktechnik und des Drucksatzes wurden durch den Offsetdruck und das Desktop Publishing¹²⁷ noch rationellere Produktionsmethoden eingeführt. Die Distributionswege wurden durch schnellere Verkehrswege, computeroptimierte Logistik und neue Vertriebskanäle – etwa das Internet – optimiert und erweitert. Die Literaturrezeption wurde durch neue Medien ergänzt, die auch vielfältige Möglichkeiten bieten, sich über Bücher zu informieren. Dennoch haben sich die fundamentalen und strukturellen Bedingungen der Schriftproduktion und -distribution bis zum Ende des 20. Jahrhunderts nicht verändert.¹²⁸

Auf dem Weg von der Einführung des Buchdrucks zum Literaturbetrieb des 19. und 20. Jahrhunderts möchte ich folgende Entwicklungen herausgreifen, weil durch sie neue Strukturen etabliert wurden, die den Umgang mit Büchern bis in unsere heutige Zeit prägen und damit ebenso Informationen übermitteln, die an das Medium Buch gekoppelt sind.

- Die Entwicklung des Buchtitels und die Etablierung bibliografischer Titelangaben.
- Die typografische Standardisierung schriftlicher Texte.
- Die Etablierung von Distributionswegen.
- Das Entstehen neuer Selektionsmechanismen bei Tradierung von Texten.

Es ist schwierig, die Aspekte einzeln zu betrachten, da sie sich wechselseitig bedingen.

¹²⁶ Stein (2006), S. 298.

¹²⁷ Auch »Computersatz«.

¹²⁸ Vgl. Stein (2006), S. 307.

Buchtitel

Als Folge der Einführung des Buchdrucks entwickelt sich das Buch vom Unikat zur Massenware. Als »Neuerung«, bezogen auf Struktur und Aufbau des Buches, bildet sich in diesem Kontext das Titelblatt heraus. Handschriftliche Kodizes hatten in den meisten Fällen kein Titelblatt. Incipit und Explizit bzw. Kolophon enthielten meist die Angaben zu Autor und Schreiber. Auch Widmungen an den Auftraggeber einer Abschrift waren hier zu finden. Man musste zumindest Anfang und Ende eines Buches lesen, um die entsprechenden Informationen zu erhalten. Der Umgang mit solchen Büchern setzte voraus, dass ein Leser oder Bibliothekar ein Buch »kannte« um gezielt darauf zurückgreifen zu können. Mit dem Buchdruck und der damit steigenden Anzahl von Büchern war es nicht mehr möglich, jedes Buch quasi »individuell« zu kennen. An diesem Punkt der Entwicklung rückten die wichtigsten Angaben auf die erste Seite des Buches. Es etablierte sich das Titelblatt. Ausschlaggebend dafür dürften in erster Linie pragmatische Aspekte beim Umgang mit Büchern gewesen sein. Als weiterer Grund kann die Nennung des Autors an prominenter Stelle gelten, die eine »Markierung des Autors« darstellte und ihn damit zu einer Art Marke machte, was dem Absatz von ihm verfasster Bücher förderlich war¹²⁹. Ebenso bot das Titelblatt dem Drucker oder Verleger, dem im Rahmen der gesteigerten Buchproduktion und eines sich entwickelnden Buchmarktes eine immer größere Bedeutung zukam, eine Gelegenheit, sich zu präsentieren.¹³⁰

Angaben über den Titel eines Buches, seinen Autor und Angaben über den Verleger und/oder Drucker sowie Ort und Jahr des Drucks rücken auf die erste Seite. Dabei entwickelte sich der Titel zu seiner heutigen Form. Ähnelte er anfangs noch der floskelhaften, beschreibenden Inhaltsangabe des Incipits, bildeten sich allmählich konkrete, kurze Buchtitel heraus.¹³¹ Die Titelangaben dienen bis heute als Metadaten zur Katalogisierung von Büchern. Auf ihnen beruhen Handelsverzeichnisse und Bibliothekskataloge. Entsprechend sind sie im Kontext unserer Buchkultur maßgeblich für den Nachweis von Texten und Textstellen, wie sie etwa im Rahmen wissenschaftlicher Arbeiten verwendet werden. Mit der Erfassung der Titelangaben für Kataloge etc. wurden zum ersten Mal standardisierte Metadaten erfasst.

Typografische Standardisierung

Mit dem Buchdruck und der zunehmenden Verbreitung von Büchern seit dem 15. Jahrhundert begann ein Prozess der Alphabetisierung und Literarisierung breiter Bevölkerungsschichten in Europa. Da immer mehr Bücher verfügbar waren, lernten immer mehr Menschen lesen, und die lesenden Menschen verlangten nach immer mehr Büchern. Die zunehmende Alphabetisierung der Bevölkerung ist somit als unmittelbare Folgeerscheinung des Buchdrucks zu sehen. Die Buchkultur hat weitgehend das Text- und Leseverständnis unserer Kultur geprägt.¹³²

Der Buchdruck hatte auch eine zunehmende Standardisierung des Schriftbildes – der Typografie – zur Folge. Zum einen wurden gedruckte Texte durch Verbesserungen der Typografie leichter lesbar; zum anderen etablierten sich typografische Standards, die eine schnelle Orientierung im Text erlauben, wie ich bereits im Abschnitt über Typografie dargestellt habe.¹³³ Diese veränderten die Möglichkeiten der Buchrezeption. Das Lesen entwickelte unterschiedliche Qualitäten: Das schnelle, »kursorische« Lesen, das einen Überblick über einen Text verschaffte und das spezialisierte, »statarische« Lesen, etwa um komplizierte Zusammenhänge erfassen zu können.

129 Assmann (1994), S. 9.

130 Schmitt, Alfred: Die Erfindung der Schrift. Rede, geh. vor d. Univ. Erlangen am 19. Juni 1938 bei d. Eröffnungsfeier d. Universitätswoche. Erlangen: Palm & Enke, 1938. (Erlanger Universitäts-Reden 22) S. 12.

131 Vgl. Stein (2006), S. 215 f.

132 Vgl. McLuhan, Marshall: Die Gutenberg-Galaxis. Das Ende des Buchzeitalters Düsseldorf/Wien: Econ, 1968. Eisenstein, Elisabeth L.: Die Druckerpresse. Kulturrevolutionen im frühen modernen Europa. Wien, New York: Springer, 1997. Giesecke, Michael: Der Buchdruck in der frühen Neuzeit. Eine historische Fallstudie über die Durchsetzung neuer Informations- und Kommunikationstechnologien. Frankfurt a. M.: Suhrkamp, 1991. – Um nur einige zu nennen.

133 Vgl. 2.1.3 Text und Gestaltung, S. 43.

Mit der Kompetenz, eine eigene Lektürewiese zu bestimmen und ein eigenes Leseverständnis zu entwickeln, wuchs auch der Bedarf nach eigenen Buchexemplaren, nicht zuletzt, um eigene Notizen oder Anmerkungen darin anbringen zu können.

Diese Weiterentwicklung des Rezeptionsverhaltens hatte ein anderes Konsumverhalten der Leseöffentlichkeit zur Folge. Von der größtenteils auf Leihverkehr und Bibliothekswesen beruhenden Literaturversorgung einer breiteren Leseöffentlichkeit im 18. Jahrhundert, der gelehrte Privatbibliotheken privilegierter Büchersammler gegenüberstanden, entwickelten sich im ausgehenden 19. Jahrhundert die Strukturen des Literaturbetriebs aus Autoren und Verlagen, Literaturkritik, Buchhandel und Bibliotheken, die sich bis heute erhalten haben. Bücher wurden immer mehr zum privaten Besitz und Gebrauch angeschafft.

Distributionswege

Handschriftlich geschriebene oder kopierte Bücher waren Unikate, die meist an einen Standort gebunden waren. Ein wertvoller handgeschriebener Kodex war keine gängige Handelsware. Auch an den Verleih war kaum zu denken. Das änderte sich, als Bücher durch die Einführung des Buchdrucks in größeren Mengen produziert werden konnten.

Zwischen der Buchproduktion, bei der man durch neue Techniken immer größere Mengen Texte drucken konnte und der immer größer werdenden Leserschaft, die durch eine gesteigerte Lesefähigkeit immer mehr Texte rezipieren konnte, stand als Bindeglied die Distribution. Die geschichtliche Entwicklung des Buchhandels ist in verschiedenen Werken ausführlich dargestellt worden.¹³⁴ Ich greife hier im Hinblick auf die Distributionsformen elektronischer Medien nur einige Aspekte heraus.

Die grundlegenden Möglichkeiten der Literaturversorgung bilden der Buchhandel und das Bibliothekswesen. Beiden liegen komplexe Strukturen zugrunde, die die Literaturversorgung des Lesepublikums gewährleisten. Der Buchhandel mit einzelnen Buchhandlungen, Fachbuchhandlungen, Grossisten und Verlagen; die Bibliotheken mit allgemeinen und wissenschaftlichen Bibliotheken, der Vernetzung in Bibliotheksverbänden, dem Fernleihverkehr etc. Der »Zugang« zu einem Buch über die entsprechenden Einrichtungen gibt einem Leser bereits Informationen über das gewünschte Buch. Ob ein Buch in der kleinen Buchhandlung »an der Ecke« oder einer Fachbuchhandlung erhältlich ist, ob und wo es dort im Regal steht oder ob es auf einem Auslagentisch liegt (oder unter dem Ladentisch gehandelt wird), kann ebenso Informationen über Inhalt, Nachfrage oder Ansehen des Buches geben wie sein Standort in einer Bibliothek. Auch der Beschaffungsweg kann indirekt Informationen über ein Buch geben. Kann es der Buchhandel über den Grossisten beziehen oder nur als Verlagsbestellung besorgen oder muss es per »Print on Demand« geordert werden? Ist es im Bestand der Bibliothek vorhanden, kann es bestellt werden oder handelt es sich um einen geschützten Bestand, der nur im Lesesaal, in wenigen Fachbibliotheken oder einem Archiv eingesehen werden kann? Der gesamte Kontext, in dem ein Buch zu haben ist, erlaubt dem, der sich mit Büchern und dem Buchhandel/Bibliothekswesen auskennt, Rückschlüsse auf das Buch: seinen Inhalt, seine Verbreitung, seine Bekanntheit etc.

Ein Buch aus einer Bibliothek bietet eventuell noch mehr Indizien, die auf alles, was mit seiner Geschichte bzw. der des enthaltenen Textes zu tun hat, schließen lassen: Der Zustand des Buches, Inventar- und Besitzstempel, Anmerkungen anderer Benutzer oder früher auch die Rückgabekarte der Bibliothek.

Neue Selektionsmechanismen der Tradierung von Texten

Mit dem Buchdruck änderten sich auch der Wert bzw. der Status eines Buches. Handgeschriebene Kodizes waren Unikate. Ihre Herstellung stand meist in einem konkreten Kontext – dem einer Klosterbibliothek, eines

¹³⁴ Vgl. zu diesem Thema: Kapp, Friedrich und Goldfriedrich, Johann: Die Geschichte des deutschen Buchhandels. 4 Bde. Leipzig: Verlag des Börsenvereins der deutschen Buchhändler, 1886–1913. Wittmann, Reinhard: Geschichte des deutschen Buchhandels. München: Beck, 1999.

Fürstenhofes oder eines Mäzens etc. Durch den Buchdruck und die massenhafte Verbreitung von Büchern verloren schriftliche Texte ihren Unikatcharakter. Die einzelne Schrift hat keinen konkreten Ort mehr. Ihre Massenproduktion und -verbreitung bedeutet einen Wertverlust des einzelnen Buches. Vervielfältigte Schriften haben viele Orte, aber keinen bestimmten oder bestimmbaren und gehen unter Umständen verloren, wenn sie nicht an ausreichend vielen Orten aufbewahrt werden.

Die Bücher können zu jedermann kommen, aber sie sind keine Einzelwesen mehr, in denen sich der abwesende Sprecher verkörpert. Ihre Dauer hängt nicht von ihrer materiellen Beschaffenheit ab, sondern von den Auswahlmechanismen neuer gesellschaftlicher Institutionen, die den Bestand des schriftlich Fixierten öffentlich präsent haben.¹³⁵

Dies bedeutet, dass mit dem Buchdruck neue Selektionsmechanismen von Texten greifen. Da das einzelne Buch an Wert verliert und nicht mehr zwingend in seinem Entstehungskontext aufbewahrt wird, gehen Texte, die nicht genügend Verbreitung finden, eher verloren.

2.1.5 Entmaterialisierung und Virtualisierung von Texten

In den obigen Abschnitten habe ich einige Beispiele dargestellt, in denen Texten durch ihre materielle Form neue Informationen und Informationsebenen zukommen, die über ihre reine Buchstabenbedeutung hinausgehen. Eine eingehende Untersuchung des Themas würde sicherlich viel weitgehendere Zusammenhänge zwischen Text und Textträger nachweisen können, als mir dies im Rahmen dieser Arbeit möglich ist. Auch wenn einige der herangezogenen Beispiele Ausnahmerecheinungen sind, sollte klargeworden sein, wie vielfältig die Möglichkeiten sind, die aus der Verbindung von Text und Papier entstehen können.

Die Verarbeitung von Texten mit Computern bedeutete zunächst eine »Entmaterialisierung« der Texte: Mit der Verbreitung der Computertechnik seit Anfang der 1980er Jahre gab es einen kurzzeitigen Einbruch im Verhältnis von Text und Textträger. Durch die Verfügbarkeit von Computern für den persönlichen Gebrauch und die Entwicklung von Textverarbeitungsprogrammen wurden Texte auf dem Computer geschrieben. Dabei dienten Computer zunächst als Ersatz für die Schreibmaschine. Die Möglichkeiten von Computern, Texte auf dem Bildschirm oder im Druck darzustellen, waren einige Zeit noch beschränkter als die von Schreibmaschinen. So war es bis Mitte der 1980er Jahre keine Selbstverständlichkeit, dass auf einem Computer deutsche Umlaute zur Verfügung standen; selten stimmte die Bildschirmansicht mit dem überein, was ein Computerdrucker nachher ausgab. Mitte der 1980er Jahre erschienen vereinzelt Bücher, deren Druckvorlagen mit dem Computer erstellt wurden. Die gestalterischen Möglichkeiten waren bei frühen Formen des Computersatzes jedoch sehr beschränkt. So gab es eine kurze Phase, in der Texte im Computer sich deutlich von Texten in Büchern unterschieden. Die technischen Einschränkungen bei der gestalteten Textdarstellung mit Hilfe von Computersystemen forcierten die Entwicklung alternativer Technologien zur Textauszeichnung.

Durch Fortschritte in der Computertechnik wie die Einführung grafischer Benutzeroberflächen fanden solche Technologien auf breiter Ebene wenig Beachtung, da die Simulation von Papier auf dem Computer den meisten Nutzern eine tiefgehende Beschäftigung mit den Hintergründen elektronischer Texte ersparte.

2.2 Texte in elektronischen Medien

Ich habe in den vorangegangenen Kapiteln die Tradition und Entwicklung schriftlicher Texte umrissen und dargestellt, dass Bücher komplexe Informationen enthalten, die über den Informationsgehalt von »aufgeschriebener Sprache« hinausgehen. Texte in Büchern haben weitere bedeutungstragende Dimensionen, die auf verschiedenen

¹³⁵ Müller, Jan-Dirk: Der Körper des Buches. Zum Medienwechsel zwischen Handschrift und Druck. In: Materialität der Kommunikation. Hg. von Hans U. Gumbrecht und K. Ludwig Pfeiffer. 2. Aufl. Frankfurt a. M.: Suhrkamp, 1995. S. 143 – 149. S. 215.

Ebenen mit dem Medium Buch verbunden sind und eine Übertragung von Texten aus Büchern in ein anderes Medium problematisch werden lassen, da sie eng an die »Techniken des Schreibens« gekoppelt sind.

Im Anschluss werde ich darstellen, welcher Techniken sich elektronische Medien bedienen, um Texte und Informationen zu speichern bzw. zu vermitteln. Zum einen sollen dabei die medialen Gemeinsamkeiten und Unterschiede zwischen elektronischen Medien und Büchern deutlich werden. Zum anderen werde ich aufzeigen, dass auch auf der Seite der elektronischen Medien bzw. in der Geschichte ihrer Entwicklung und Etablierung Tendenzen erkennbar sind, die die Entwicklung der eingangs dargestellten verschiedenen Konzepte elektronischer Editionen¹³⁶ begünstigt haben.

Um die Tradition und Entwicklung elektronischer Texte anschaulich darzustellen, greife ich im Folgenden auf eine – zugegeben teilweise etwas konstruierte – Analogie zur gesprochenen Sprache und ihrem Verhältnis zu geschriebenen Texten zurück.

2.2.1 Speicherung von Texten in elektronischen Medien

Ähnlich, wie gesprochene Texte aus einer Abfolge von Lauten bestehen, besteht ein elektronischer Text, wie nahezu alle elektronischen Daten, aus einer Abfolge elektrischer Impulse. Während ein sprachlicher Text in Form von Schallwellen durch die Luft übertragen wird, wird ein elektronischer Text in Form von elektrischen Impulsen, die durch einen Leiter fließen, übertragen. Wie sich bei gesprochener Sprache der Sinn, also ihr Informationsgehalt, durch die Abfolge der Laute in der Zeit ergibt, ergibt sich der Informationsgehalt der elektrischen Impulse ebenfalls aus ihrer Abfolge.¹³⁷ Wie Schallwellen verhalten, nachdem ein Text gesprochen worden ist und die Laute ihr Trägermedium, die Luft, durchlaufen haben, ist auch die Impulsfolge eines elektronischen Textes nicht mehr vorhanden, wenn sie ihr Trägermedium, etwa ein Kabel, durchlaufen hat.

Um elektronische Texte zu bewahren, müssen sie in einer geeigneten Form gespeichert werden, beispielsweise in einem Speicherbaustein. Speicherbausteine kann man als das »Gedächtnis« eines Computers betrachten, das ähnlich wie das Gedächtnis eines Menschen funktioniert. Wie das menschliche Gehirn gehörte Texte vergessen kann, können auch elektronische Speicherbausteine Texte »vergessen«. Im Arbeitsspeicher eines Computers gehen Texte verloren, wenn man den Strom abschaltet, eine Erfahrung, die man macht, wenn man den Computer ausgeschaltet hat, ohne einen gerade geschriebenen Text vorher auf einem geeigneten dauerhaften Speichermedium, beispielsweise einer Diskette oder der Festplatte des Computers, abzuspeichern. Genauso, wie man einen Satz, den man nicht vergessen will, auf einem Zettel notiert, benötigt man für elektronische Texte einen materiellen Träger, um sie dauerhaft zu speichern. Was bei gesprochener Sprache das Papier ist, auf dem man den Text notiert, ist bei elektronischen Medien die Festplatte, die Diskette oder die CD-ROM,¹³⁸ auf deren Oberflächen die elektrischen Impulse magnetisch oder optisch aufgezeichnet werden.

2.2.2 Datenformat und Darstellungsformat

Um die auf einem materiellen Träger gespeicherten Daten – einen auf die Festplatte oder eine CD-ROM geschriebenen elektronischen Text – im Computer verarbeiten zu können, etwa, um den Text auf dem Bildschirm darzustellen und zu bearbeiten oder um ihn auszudrucken, müssen die Daten erst wieder vom Speichermedium in den Computer eingelesen, also wieder in elektrische Impulsfolgen umgewandelt werden. Damit die Daten nach dem Schreiben wieder gelesen werden können, müssen sie in einer geeigneten Form aufgeschrieben wer-

¹³⁶ Vgl. 1.1 Typen elektronischer Editionen, S. 9.

¹³⁷ Vgl. Grundlagen der Digitaltechnik, S. 71.

¹³⁸ Ich nenne hier nur einige derzeit gängige Speichermedien. Für andere Speichermedien – etwa DVD, Flash-Speicher etc. – gilt dies natürlich ebenso.

den. Ein elektronischer Text benötigt zur dauerhaften Speicherung eine geeignete Form bzw. ein Format. Im Fall von elektronischen Daten spricht man von einem Datenformat.¹³⁹

Elektronische Texte haben demnach mindestens zwei Formate: ein Speicher- bzw. Datenformat, in dem sie längerfristig gespeichert werden können und ein Darstellungsformat, in dem der elektronische Text später auf einem geeigneten Ausgabegerät für einen Menschen lesbar angezeigt wird. Damit ein Leser elektronische Texte lesen kann, müssen elektronische Texte vom Datenformat in ein Darstellungsformat gebracht werden: Die in einem Datenformat auf einem Datenträger abgelegten Informationen werden wieder in eine Form umgesetzt, die der Computer verarbeiten kann; der Computer erzeugt ein geeignetes Darstellungsformat, das den Text in einer für den Menschen lesbaren Form anzeigt.

2.2.3 Kodierung elektronischer Texte

Damit elektronischen Texten neben dem Text aus Buchstaben weitere Informationen mitgegeben werden können, etwa darüber, wie der Text darzustellen ist, müssen diese Informationen in einer bestimmten Form, die der Computer auswerten kann, eingeschrieben – kodiert – werden. Wichtig für die elektronische Verarbeitung von Texten ist, dass durch die Kodierung gewährleistet wird, dass Informationen über den Text von den eigentlichen Textinformationen getrennt werden. Damit ein Computer die Daten richtig auswerten kann, muss eindeutig sein, ob es sich bei den Informationen »Franklin« und »fett« um Angaben über die Darstellung des Textes handelt, der in der Schriftart Franklin in einem fetten Schriftschnitt dargestellt werden soll oder um den Namen und die körperliche Erscheinung einer im Text beschriebenen Person. Was für einen Leser schriftlicher Texte evident erscheint, ist für die Datenverarbeitung im Computer von Bedeutung. Kein Leser würde einen fetten Schriftschnitt mit dem Adjektiv »fett«, wie es in einem Text vorkommen kann, in Verbindung bringen. Die Information fett ist jeweils auf einer völlig anderen Ebene eines gedruckten Textes angesiedelt – einmal auf der Ebene der Buchstaben »f« »e« »t« »t« und einmal auf der Ebene der Typografie. Bei der Informationsverarbeitung im Computer werden aber letztendlich alle Informationen auf reine Zeichenkodes reduziert.¹⁴⁰ Auf dieser Ebene ist es wichtig, das eindeutig unterschieden werden kann, ob »fett« nun in den Text gehört und als Wort im Text dargestellt wird oder ob es sich um eine Formatierungsangabe handelt, die besagt, dass der entsprechende Text fett dargestellt wird. Die Kodierung muss gewährleisten, dass solche Informationen nicht mit den eigentlichen Textinformationen durcheinandergeraten.

Aus der Art der Kodierung ergibt sich das Datenformat, in dem die elektronischen Texte gespeichert werden.

2.2.4 Datenformate für formatierte Texte

Die Formatierungen schriftlicher Texte haben meist Bedeutungen, die dem Leser bekannt oder zumindest offensichtlich sind. Den gedanklichen Prozess, der den Informationsgehalt von Formatierungen auswertet, macht sich ein Leser selten bewusst. Er gehört zur allgemeinen Lesekompetenz. Auch wenn die Art der Formatierung nicht in allen Texten gleich ist, erkennt ein Leser die wichtigsten Strukturen eines Textes meist auf den ersten Blick.¹⁴¹

Pure elektronische Texte, die lediglich aus den im Text enthaltenen Buchstaben bestehen,¹⁴² enthalten keine Informationen über ihre Formatierung – mit Ausnahme von Zeilen- oder Absatzumbrüchen, wenn man diese zu den Formatierungen zählt. Sie enthalten weder genauere Angaben zur Darstellung, noch weitere Informationen,

139 Diese Darstellung ist – technisch gesehen – stark vereinfacht. Die Daten erfahren unterschiedlichste Umwandlungsschritte, die hier im Einzelnen aber keine Rolle spielen, da ein Computerbenutzer im Normalfall nicht mit ihnen in Berührung kommt. Datenformate und Darstellungsformate sind jedoch für normale Computernutzer von Bedeutung.

140 Vgl. Grundlagen der Digitaltechnik, S. 71.

141 Siehe 2.1.3 Text und Gestaltung, S. 43.

142 Z. B. reine Zeichendaten, von Word auch als »nur Text« bekannt.

wie sie sich direkt oder indirekt aus Büchern ablesen lassen. Mit solchen elektronischen Texten hatte man in der Frühzeit der Computertechnik hauptsächlich zu tun.

Die ersten Datenformate für elektronische Texte, die Informationen zur Textformatierung beinhalteten und die sich auf einer breiteren Ebene durchsetzten, waren Datenformate von Textverarbeitungsprogrammen. Die Möglichkeiten zur formatierten Textdarstellung waren damals noch stark begrenzt – sowohl bei der Bildschirmdarstellung als auch beim Ausdruck auf den damals verbreiteten Matrixdruckern. Als das heute wohl am weitesten verbreitete Textverarbeitungsprogramm Microsoft Word 1983 in der Version 1.0 auf den Markt kam, war es anderen Textverarbeitungsprogrammen, wie etwa den damals aktuellen 3.0 Versionen von WordPerfect oder WordStar, dadurch überlegen, dass es die Möglichkeit bot, fette und kursive Schriftschnitte auf dem Bildschirm darzustellen – allerdings nur in der Systemschriftart. Die anderen Programme konnten Text, der fett oder kursiv ausgedruckt werden sollte, auf dem Bildschirm nur durch eine andere Farbe der Standard-Bildschirmschrift oder mit Hilfe von Steuerzeichen kennzeichnen.¹⁴³ Da dies für den Anwender nicht sehr komfortabel war, konzentrierte sich die Entwicklung darauf, formatierten Text auf dem Bildschirm darstellbar zu machen. Der Durchbruch gelang dabei mit der zunehmenden Verbreitung von grafischen Benutzeroberflächen wie Microsoft Windows oder der des Apple Macintosh Ende der 1980er Jahre.

War die Darstellung und Speicherung formatierter Texte auf einem Computer bald kein Problem mehr, ergaben sich neue Probleme beim Austausch formatierter Texte zwischen verschiedenen Computern. Eine einheitliche Textdarstellung konnte selbst zwischen Computern, die mit dem gleichen Betriebssystem arbeiteten, problematisch sein. Die Gründe waren sowohl in der Hardware zu suchen – etwa bei unterschiedlichen Grafikkarten – oder sie konnten durch die Software bedingt sein, etwa, weil auf den Rechnern unterschiedliche Schriftarten oder Druckertreiber installiert waren. Der Austausch elektronischer Texte zwischen Computern mit verschiedenen Betriebssystemen scheiterte teilweise schon auf der Ebene der reinen Zeichendaten.¹⁴⁴ Noch komplexer wurde der Austausch formatierter Texte, da auf den jeweiligen Rechnern neben den gleichen Zeichensätzen auch die gleichen Schriftarten (z. B. Timens New Roman, Helvetika, Arial) vorhanden sein mussten. Die einzelnen Betriebssysteme sind hier eigene Wege gegangen. Vor dem Hintergrund der zunehmenden Vernetzung unterschiedlicher Computer, etwa über das Internet, etablieren sich zurzeit neue, einheitliche Standards. Auf der Ebene der Zeichenkodierung setzt sich etwa *UNICODE* als weltweiter, systemübergreifender Standard durch; bei den Schriftarten stehen mit den Open Type Fonts einheitliche Schriftarten für die meisten Computersysteme zur Verfügung.

Neben den technischen Problemen gibt es weitere, beispielsweise lizenzrechtliche. Damit Computer elektronische Texte korrekt formatiert darstellen können, müssen sie auf die jeweilige Schriftart, den so genannten *Font*, mit dem der Text erstellt wurde, zurückgreifen können. Dieser muss entweder auf dem Computer, auf dem der Text dargestellt werden soll, vorhanden sein oder er muss mit dem Text mitgeliefert werden. Im ersten Fall muss derjenige, auf dessen Computer sich kein passender *Font* findet, diesen auf seinem System installieren, wozu er ihn ggf. noch kaufen muss. Im zweiten Fall bekommt der Verfasser bzw. Vertreiber des entsprechenden Textes unter Umständen Probleme mit der Firma, die die Schrift vertreibt, da er die Schriftart unrechtmäßig verbreitet und sich somit der Softwarepiraterie schuldig macht.

Dateiformate, die Textformatierungen zuverlässig speichern und übertragen konnten, waren zuerst nur teuren Spezialprogrammen, etwa Satzprogrammen wie QuarkXPress, das 1987 zuerst nur für den Macintosh

143 Mit einer damals gängigen »Hercules«-Grafikkarte konnten 80 Zeichen in 25 Zeilen in (meist) grüner Systemschrift auf schwarzem Hintergrund dargestellt werden. Dass Text fett formatiert ausgedruckt werden sollte, stellte WordPerfect auf dem Bildschirm etwa so dar:

fett ▶ **fetter Text** ◀ **fett**

144 Auch wenn beinahe alle Computersysteme den gleichen Standard-ASCII-Zeichencode (7 Bit ASCII) verwenden, ist dies bei Erweiterungen auf 8-Bit ASCII nicht mehr der Fall. Hier arbeiten DOS, Windows, Macintosh und Unix mit 8 Bit ASCII, ANSI, IBM PC, ISO 646, MS-DOS, ISO 8859 etc. mit unterschiedlichen Standards.

verfügbar war, vorbehalten. Bereits seit 1984 gab es die von der Firma Adobe entwickelte Seitenbeschreibungssprache *PostScript*, die allerdings eher eine Programmiersprache ist und ursprünglich zur Ansteuerung von Belichtungsmaschinen für Druckvorlagen im Offsetdruck genutzt wurde. Eine adäquate Darstellung auf dem Bildschirm war lange Zeit problematisch und aus lizenzrechtlichen Gründen – Adobe hat sich alles, was mit *Post-Script* zusammen hing, gut bezahlen lassen – sehr teuer und blieb somit Expertensystemen vorbehalten.

Das erste Dateiformat, das allgemein Verbreitung fand und Formatsicherheit bot, war das 1993 von Adobe eingeführte *Portable Document Format*. Programme zur Erstellung und zur Anzeige von Texten im *PDF* sind für verschiedene Computersysteme verfügbar und ermöglichen es, Dokumente speziell für die Verwendung in unterschiedlichen Medien einzurichten. Damit ist das *PDF* sehr vielseitig und kann sowohl für die Bildschirmdarstellung als auch für die Druckvorstufe verwendet werden. Es ist das zurzeit bevorzugte Dateiformat bei buchnahen elektronischen Editionen. Zum einen ermöglicht es, auch komplexe Formatierungen zuverlässig zu speichern und wieder darzustellen; zum anderen können Dokumente im *PDF* für elektronische Medien z. B. mit *Hypertext*funktionen versehen werden. Hinzu kommt, dass unterschiedliche Versionen eines Dokuments – eine für die Druckvorstufe und eine für die elektronische Nutzung eingerichtete, nahezu in einem Arbeitsgang erstellt werden können.

Datenformate, die darauf ausgelegt sind, Texte unter möglichst genauer Beibehaltung ihres Darstellungsformats zu speichern, haben einige gravierende Nachteile:

Informationen, die im Text lediglich »typografisch kodiert« sind, sich also dem Leser aufgrund seiner Kompetenz im Umgang mit Texten erschließen, können nicht elektronisch ausgewertet werden. Sie stehen damit nicht für einen mediengerechten Umgang mit elektronischen Texten – etwa zur Realisierung komplexer *Retrieval*funktionen – zur Verfügung. Sind etwa in einem Drama, das als elektronischer Text vorliegt, die Sprecher fett formatiert, erkennt ein Leser die Struktur sofort. Er sieht vielleicht aufgrund der hervorgehobenen Sprecher zuerst, dass es sich überhaupt um ein Drama handelt. Ein Leser kann gezielt die Äußerungen einer bestimmten Figur des Dramas nach bestimmten Wörtern durchsuchen. Ein Computer kann dagegen nicht automatisch unterscheiden, ob der fett dargestellte Text nun einen Sprecher kennzeichnet oder ob es sich etwa um eine Überschrift handelt. Noch problematischer sind Informationen, die ein komplexeres Hintergrundwissen voraussetzen, wie etwa das oben genannte Beispiel mit dem Begriff »Reich«. ¹⁴⁵ Solche Informationen können bei formatierungsbezogenen Datenformaten überhaupt nicht elektronisch ausgewertet werden.

Wie ich dargestellt habe, ist es unter Umständen problematisch, Formatierungen bei der Speicherung und beim Austausch von Texten beizubehalten. Ist aus technischen Gründen die formatierte Darstellung nicht mehr oder nur eingeschränkt möglich, können auch die mit der Formatierung verbundenen Informationen verloren gehen. Schlimmstenfalls können Daten, die in einem Format abgelegt wurden, das später nicht mehr unterstützt wird, verloren sein. ¹⁴⁶

Elektronischen Texten geht, aufgrund ihres elektronischen Mediums und der damit verbundenen Präsentations- und Vertriebsformen, ein großer Teil der kontextbezogenen Informationen, die Bücher mit sich bringen können, verloren. Viele dieser Informationen sind bei Büchern in Form von »Indizien« abzulesen. So kann etwa der Einband eines Buches Rückschlüsse auf sein Alter und seine Nutzung zulassen. ¹⁴⁷ Auch Aspekte

¹⁴⁵ Vgl. S. 46.

¹⁴⁶ Als Beispiel aus der Praxis kann ich einen Fall anführen, in dem es notwendig war, die ersten drei Bände eines Kunstlexikons auf Grundlage der gedruckten Bände neu zu digitalisieren, da die vorhandenen Satzdateien in einem frühen QuarkXPress-Datenformat nicht mehr gelesen werden konnten. Aktuelle Versionen dieses Programms unterstützten das Datenformat nicht mehr, alte Versionen des Programms standen genauso wenig zur Verfügung wie Rechner, auf denen eine alte Programmversion hätte laufen können.

¹⁴⁷ Vgl. 2.1.4 Buchproduktion und Distribution, S. 55.

der Distribution stehen bei elektronischen Texten seltener zur Verfügung. Bei einem Text, der »gegoogelt«¹⁴⁸ wurde, ist nicht immer ersichtlich, aus welchem Kontext er stammt.

2.2.5 Semantische Kodierung

Alternativ zur Kodierung von Formatierungen gibt es die Möglichkeit der semantischen Kodierung. Die Grundidee der semantischen Kodierung ist die Trennung von Form und Inhalt. Sie ist eine der wichtigsten Voraussetzungen, um Text elektronisch sinnvoll verarbeiten zu können.

Die Form schriftlicher Texte ist, wie ich dargestellt habe, in den meisten Fällen relevant für den Inhalt. Ist in einem schriftlichen Text etwa ein Wort kursiv gedruckt, kann dies, je nach Kontext, unterschiedliche Bedeutungen haben. Auf jeden Fall ist es eine Hervorhebung des Wortes. Sie kann einfach nur ein besonders wichtiges Wort, etwa einen Schlüsselbegriff, im Text hervorheben. Es kann auch sein, dass es sich um einen Verweis handelt und der kursiv gedruckte Begriff an anderer Stelle erläutert wird. Genauso können Kursivierungen verwendet werden, um beispielsweise Titel von Büchern, Filmen oder Opern etc. zu kennzeichnen oder auch um fremdsprachigen Text, etwa ein lateinisches Wort, zu markieren. Ich bezeichne diese Form der Kodierung als »typografische Kodierung«.

Das Konzept der semantischen Kodierung beruht darauf, einen Text nicht seiner späteren Formatierung gemäß zu kodieren, sondern die Kodierung zu benutzen, um unmittelbar die Bedeutung eines bestimmten Textbereiches zu beschreiben. Anstatt eine Textstelle als »kursiv« zu kodieren, kann man sie auch als »Schlüsselwort« kodieren; anstatt Informationen wie »Schriftart Helvetika, Schriftgröße 16 Punkt, Textausrichtung zentriert«, anzugeben, kann man etwa auch »Überschrift erster Ordnung« oder »Kapitelüberschrift« verwenden. Anstelle »fett« zur Hervorhebung der Sprecher in einem Drama zu verwenden, kann man diese gleich als »Sprecher« kennzeichnen. Eine formale Kodierung stellt im Endeffekt eine doppelte Kodierung dar: Die Formatierung eines schriftlichen Textes ist in den meisten Fällen schon eine Kodierung von Informationen, die über die buchstäbliche Textinformation hinausgeht. Eine formatierungsorientierte Kodierung kodiert diesen (typografischen) Kode nochmals. Da elektronische Datenformate diese Möglichkeit bieten, ist es sinnvoll, hier nach Möglichkeit direkt die ursprüngliche Information anzugeben, die mit der Formatierung übermittelt werden sollte.

Die Trennung von Form und Inhalt ermöglicht es, inhaltsbezogene Informationen unmittelbar mit dem Text zu speichern, während die Form bzw. die Darstellung von einer eigenen Instanz gesteuert wird, die sich nach der inhaltlichen Kodierung richtet.

Die Grundlagen der semantischen Kodierung stammen aus der Zeit, als elektronische Texte hauptsächlich aus *ASCII*-Zeichen bestanden. Ende der 1960er Jahre wurde das Konzept von *Auszeichnungssprachen* entwickelt, die auf eine inhaltliche Auszeichnung von Texten setzten.¹⁴⁹ Diese Konzepte wurden damals hauptsächlich von Experten zur Kenntnis genommen.¹⁵⁰ Die allgemeine Entwicklung ging eher in die Richtung, Möglichkeiten zu schaffen, Formatierungen elektronischer Texte darzustellen und mit den Texten zu speichern. Aus heutiger Sicht könnte man dabei beinahe von einer Fehlentwicklung sprechen. So war das Konzept der 1969 von IBM vorgestellten Beschreibungssprache Generalized Markup Language (GML) elektronischen Medien angemessener als das Festhalten an starren Formatierungen, wie man sie vom Umgang mit schriftlichen Texten her gewohnt war.

148 Die Verwendung der Internet-Suchmaschine Google gehört mittlerweile zu den Standardpraktiken der elektronischen Informationsbeschaffung, siehe <http://www.google.de> (18.05.2008).

149 Die Idee, Formate und Informationen in elektronischen Dokumenten zu trennen, wird allgemein auf William W. Tunnicliffe zurückgeführt, der seine Idee des »generic coding« 1967 veröffentlichte.

150 Um eine Orientierung an der technischen Entwicklung zu ermöglichen: Der erste serienmäßig produzierte Mikroprozessor wurde 1971 von Intel hergestellt. Die Grundidee von Auszeichnungssprachen reicht also in eine Zeit zurück, als Computer noch mit mechanischen Relais, Röhren oder – damals ganz modern – mit Transistoren arbeiteten.

Anfang der 1990er Jahre wurde das *World Wide Web* populär. Die wichtigste Grundlage dafür, dass das bisher hauptsächlich militärisch und wissenschaftlich genutzte Internet¹⁵¹ für breitere Nutzerschichten interessant wurde, war die Entwicklung des ersten grafikfähigen Webbrowsers Mosaic im Jahr 1993. Zur Realisierung der *Hypertextfunktionen* setzte das *WWW* auf eine *Auszeichnungssprache* zur Kodierung von Texten, die *Hyper Text Markup Language, HTML*. *HTML* wurde auf Basis der *Standard Generalized Markup Language* entwickelt, die wiederum eine Weiterentwicklung der *Generalized Markup Language GML* (s. o.) ist. *HTML* erlaubte die semantische Kodierung von grundlegenden Textstrukturen, etwa Überschriften verschiedener Hierarchieebenen, Absätzen, Zitaten oder Autoren, sah aber daneben auch rein formatierungsbezogene Kodierungen, also etwa fett und kursiv, vor.

Im weltumspannenden Internet mit seinem technischen Hintergrund setzte man damit tendenziell auf die semantische Kodierung von Texten, während der Schwerpunkt bei den Einzelplatzrechnern, den Personal Computern (PCs), auf der Formatierung von Texten lag. Neben der Buchtradition ist dies sicher ein weiterer Grund für die beiden Typen elektronischer Editionen: Am Buch orientierte elektronische Editionen sind, wie Bücher, eher für die Benutzung am einzelnen Arbeitsplatz konzipiert; elektronischen Medien nahe Editionen für die Verbreitung über das Internet und somit für die weltweite Nutzung gedacht.

Mit der zunehmenden Verfügbarkeit breitbandiger Internetverbindungen in den letzten 10 Jahren werden immer mehr Einzelplatzrechner vernetzt. Die beiden ehemals getrennten Computerwelten, die der weltweit vernetzten Großrechner und die der persönlich genutzten Einzelplatzrechner, wachsen zusammen. Damit entsteht das Bedürfnis, auch die beiden Welten elektronischer Textkodierung zu verbinden. Das oben erwähnte *PDF* stellt einen Schritt in diese Richtung dar. Neben für die Druckvorstufe eingerichteten *PDF*-Dokumenten können auch für die Übertragung im Internet optimierte Versionen erstellt werden. Daneben ermöglicht die zum Betrachten eines *PDF* benötigte Reader-Software immer mehr elektronische Funktionen wie *Hyperlinks* oder Volltextsuche. Umfangreiche *Retrievalfunktionen* bleiben aber spezialisierten Programmen vorbehalten.

2.2.6 X-Media-Publishing

Sowohl auf Seiten der Textproduktion als auch der Rezeption ist der Einsatz elektronischer Hilfsmittel und Medien nicht mehr wegzudenken. Im Gegenteil: Texte und ihre elektronische Verarbeitung rücken immer dichter zusammen. Dies gilt nicht nur für das Schreiben von Texten mit Hilfe eines Textverarbeitungsprogramms oder das Lesen von Texten im Internet. Vielmehr kommen beim Erstellen und Verwalten großer Textmengen, etwa bei Zeitungs- und Zeitschriftenverlagen oder im Bereich der technischen Dokumentation, immer mehr elektronische Textsysteme – Redaktionssysteme, Content Management-Systeme – zum Einsatz, mit deren Hilfe Texte geschrieben, gespeichert und zur Publikation zusammengestellt werden. Dabei werden verschiedene Medien, etwa Druck und Internet, nebeneinander bedient. Die Texte werden über komplexe *Retrievalfunktionen* in Datenbanken verfügbar gemacht; sie können verlinkt und mit gängigen Technologien elektronischer Medien ausgewertet werden. Daneben können sie für die Ausgabe in unterschiedlichen Medien aufbereitet werden. Geeignete Datenformate für solche komplexen Textumgebungen beruhen auf dem Prinzip der semantischen Kodierung. Mit Hilfe geeigneter Kodierungsmodelle werden so genannte X-Media Publishing Konzepte (cross media publishing; medienübergreifende Publikation; auch »multiple media publishing« oder »single source publishing«), die Publikation gleicher Inhalte in verschiedenen Formen und Medien (Internet, Datenbanken, CD-ROM, eBook, Druck etc.), realisiert. Auf der Grundlage semantischer Kodierung wird ein universelles Speicher- und Austauschformat für die hier erforderliche medienneutrale Datenvorhaltung verwirklicht. Um nicht für jedes Medium oder jede

151 Das Internet selber hat eine komplexe Entwicklungsgeschichte hinter sich. Es ging 1969 aus dem seit 1962 eingerichteten ARPANET der Advanced Research Project Agency (ARPA) des US-Verteidigungsministeriums hervor. Der Begriff Internet wird seit 1982 verwendet, seitdem die Regeln für den Datenaustausch nach dem TCP/IP Protokoll (Transmission Control Protocol/Internet Protocol) geregelt werden.

Verwendung einen eigenen, an die jeweiligen technischen Voraussetzungen angepassten Text vorhalten zu müssen – dies führt bei Veränderungen des zugrunde liegenden Textes dazu, dass diese Veränderungen jeweils an allen Ausgangstexten vorgenommen werden müssen –, greift man beim Konzept des X-Media Publishings auf eine einheitliche Textgrundlage zurück. Diese Entwicklung ermöglicht es, die Verknüpfungen von Texten, Apparaten und Materialien sowohl in Form von *Hyperlinks* zu realisieren als auch herkömmliche Drucklayouts mit einer Verknüpfung über Lemmata oder Fuß-/Endnotenzeichen zu erstellen.

2.3 Elektronische Editionen und Editionswissenschaft

In Anbetracht der Möglichkeiten und Veränderungen, die elektronische Medien für den Umgang mit Texten bringen, wurde bereits die Frage nach der Notwendigkeit einer neuen Editionswissenschaft aufgeworfen.¹⁵² Allerdings wurde diese Frage vornehmlich vor dem Hintergrund von *Hypertext*-Technologien diskutiert und in diesem Kontext im Wesentlichen verneint. Der Konsens der Diskussion lautete etwa, dass *Hypertexte* als die Umsetzung konventioneller Nachweis- und Verweistechiken mit elektronischen Mitteln betrachtet werden könnten.¹⁵³

Genauso wie elektronische Medien nicht das Ende des Buches herbeigeführt haben, wie es mit dem Aufkommen der elektronischen Medien Anfang der 1990er Jahre oftmals prognostiziert wurde, wird auch eine auf Bücher bezogene Editionswissenschaft weiterhin Bestand haben. Zum einen liegen noch die meisten Quellen und Textzeugen auf Papier vor, genauso wie die Mehrheit aktueller Editionsprojekte noch als Buchausgaben ausgelegt sind. Ein Großteil der wissenschaftlichen Editionsarbeit, ihre Disziplinen und deren Methoden sind überhaupt nicht von elektronischen Medien betroffen oder benutzen sie nur als Hilfsmittel. Zum anderen muss sich noch erweisen, ob die Kultur und Technik elektronischer Medien so dauerhaft ist, dass sie als Gegenstand editionswissenschaftlicher Arbeit eine größere Bedeutung gewinnen kann. Besonders im Vergleich der Traditionen der – zumindest seit dem Mittelalter bestehenden – Buchkultur mit der kurzen Spanne, in der elektronische Medien im Literaturbetrieb eine Rolle spielen, ist die Forderung, eine komplette wissenschaftliche Disziplin umzustrukturieren, sicher nicht gerechtfertigt.

Dennoch muss die Frage erlaubt sein, inwieweit sich eine zukünftige Editionswissenschaft den Anforderungen elektronischer Medien stellen kann. Schließlich haben elektronische Medien innerhalb eines relativ kurzen Zeitraums bedeutende Veränderungen in der Kommunikationsstruktur der Gesellschaft bewirkt und die Produktion wie die Rezeption von Texten beeinflusst. Für eine Beschäftigung der Editionswissenschaft mit elektronischen Medien sehe ich zurzeit drei wesentliche Ansatzpunkte:

- Die Nutzbarmachung bzw. Einrichtung vorhandener Editionen für elektronische Medien.
- Die Erarbeitung neuer, zukunftssicherer Editionen, die aktuelle und zukünftige Publikationsmedien bedienen können.
- Die Entwicklung von Konzepten, die Entstehung und Entwicklung von Texten in elektronischen Medien, also etwa die Arbeit eines Autors an seinem Computer, dokumentieren.

152 Kamzelak, Roland: *Hypermedia – brauchen wir eine neue Editionswissenschaft?* In: *Computergestützte Text-Edition*. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu *Editio* 12) S. 117-126.

153 Vgl. auch Gabler, Hans-Walter: *Rezension: Roland Kamzelak (Hg.): Computergestützte Text-Edition*. <http://computerphilologie.uni-muenchen.de/jg01/gabler.html> (18.05.2008).

2.3.1 Einrichtung vorhandener Editionen für elektronische Medien

Die Einrichtung vorhandener Bucheditionen für elektronische Medien habe ich bereits im Kapitel über elektronischen Medien nahe stehende elektronische Editionen angesprochen.¹⁵⁴ Etablierte Editionen werden meist durch Retrodigitalisierung für elektronische Medien eingerichtet. Hier liegen einige Erfahrungswerte vor. Bei den bisher verbreiteten Projekten lassen sich zwei Probleme beobachten:

- Durch die Verwendung eigener, proprietärer Datenformate können die Daten unbrauchbar werden, etwa wenn eine Firma die zum Lesen der Daten notwendige Software nicht mehr pflegt.
- Die Bindung an bestimmte Speichermedien macht die Zukunftssicherheit einer elektronischen Edition von der Alterungsbeständigkeit des Datenträgers abhängig.

Verlage, die in größerem Umfang elektronische Editionen herausgeben, greifen gerne auf eigene Softwarelösungen zurück. Dies hat den Vorteil, dass das angestrebte Editions-konzept optimal umgesetzt werden kann. Weitere Vorteile für Verlage liegen darin, dass sie hierbei ihre Leistungsschutzvorstellungen am besten realisieren können, etwa, indem sie die Texte in einer kodierten Form verkaufen, die nur mit der entsprechenden Software gelesen werden kann. In Verbindung mit der Möglichkeit, Funktionen zum Textexport einzuschränken, kann einer freien Verbreitung der elektronischen Texte entgegengewirkt werden.

Beispiele für spezialisierte Eigenentwicklungen sind etwa die Software der DIGITALEN BIBLIOTHEK des Verlags Directmedia,¹⁵⁵ die ViewLit Software der Firma Carsten Worm Infosoftware¹⁵⁶ oder die Datenbanken der Firma Chatwyck Healey.¹⁵⁷ Eigene Lösungen, sowohl für CD-ROM als auch für Online-Medien hat auch der Verlag Text und Kritik entwickelt, etwa für das KRITISCHE LEXIKON ZUR DEUTSCHSPRACHIGEN GEGENWARTSLITERATUR – KLG.¹⁵⁸

Die Textdaten stehen bei den meisten Programmen nicht in einem proprietären Datenformat zur Verfügung, sondern sind lediglich mittels der speziell darauf abgestimmten Software nutzbar. Hier machen sich Entwicklungen auf dem Gebiet der Computertechnik bzw. der Software leider negativ bemerkbar, da in den seltensten Fällen auf eine ausreichende Abwärtskompatibilität geachtet wird. Erfahrungsgemäß lässt sich eine Software, die nicht gepflegt und weiterentwickelt wird, bereits zwei Computergenerationen später nicht mehr auf aktuellen Rechnersystemen ausführen. Daten, die zur Benutzung mittels dieser Software eingerichtet wurden, sind damit nicht mehr zugänglich.

Als ehemals prominentes Beispiel möchte ich hier das Programm WordCruncher anführen. Das Ende der 1980er Jahre ursprünglich unter dem Namen »Brigham Young Concordance Program (BYC)« in den USA entwickelte und von der Firma Electronic Text Corporation vertriebene Programm WordCruncher war eines der ersten diesbezüglichen Programme, das Verbreitung fand.¹⁵⁹ Ursprünglich für das Betriebssystem DOS entwickelt, wurden bis zur Jahrtausendwende auch einige Windows-Versionen vertrieben.

WordCruncher besteht aus zwei Programmen: WC-Index zum Erstellen der zur Realisierung der Funktionen erforderlichen Indexliste und WC-View zur Textdarstellung und Recherche. WC-Index erstellte auf der Grundlage einer ASCII-Textdatei einen Wortindex. Dabei konnten verschiedene ID-Level vergeben werden, die die Struktur des Dokuments – Bücher, Abschnitte, Kapitel usw. – abbildeten. Jedes einzelne Wort wurde in den Index auf-

154 Vgl. 1.4 Elektronischen Medien nahe stehende elektronische Editionen. S. 17.

155 <http://www.digitale-bibliothek.de/> (18.05.2008).

156 <http://www.infosoftware.de/> (18.05.2008).

157 Vgl. <http://kafka.chadwyck.com/> (18.05.2008). <http://goethe.chadwyck.com/> (18.05.2008). <http://schiller.chadwyck.com/> (18.05.2008). <http://brecht.chadwyck.com/> (18.05.2008).

158 Kritisches Lexikon zur deutschsprachigen Gegenwartsliteratur. Hg. von Heinz Ludwig Arnold. München: Edition Text + Kritik. Als Loseblattsammlung, auf CD-ROM und online. <http://www.klgonline.de> (18.05.2008).

159 Vgl. Bradley, John und Rockwell, Geoffrey: TACT Web 1.0 Workbook: <http://tactweb.humanities.mcmaster.ca/tactweb/doc/twintro.htm> (18.05.2008). Eremen, Harald: Suchen, finden, lesen. WordCruncher – Text- und Retrieval Software. In: ct, 3/1991, S. 152 ff.; WordCruncher: <http://wordcruncher.byu.edu/wordcruncher/default.htm> (18.05.2008).

genommen. Daneben erlaubte WC-Index das Aufstellen so genannter Stopp-Wort-Listen. Stopp-Wort-Listen enthalten Wörter, die nicht in den Index aufgenommen werden sollen, etwa Artikel, Pronomen etc., die häufig vorkommen, meist jedoch nicht gesucht werden. WC-View diente einerseits zur Darstellung der Texte und ermöglichte schon weitentwickelte *Retrieval*-Funktionen, etwa die Verwendung von Platzhalterzeichen oder die Suche von Wortkombinationen, wobei die Suchumgebung frei eingeschränkt werden konnte. Ergebnisse wurden als Fundstellenliste mit Kontext und Ort, der sich aus den Level-Einträgen ergab, ausgegeben. Daneben konnte der Benutzer Lesezeichen setzen. Eine Besonderheit von WordCruncher war die Möglichkeit, Konkordanzlisten zu erstellen und auszugeben. Aufgrund dieser Möglichkeit wurde er gern für linguistische Auswertungen von Texten benutzt.

Neben diversen linguistischen Projekten wurde WordCruncher unter anderem für die Erstellung des WORDCRUNCHER-COMPUTERBOOK DER VORLESUNGEN KANTS der Arbeitsstelle Kant-Ausgabe der Göttinger Akademie der Wissenschaften¹⁶⁰ und die CD-ROM ROBERT MUSIL: DER LITERARISCHE NACHLASS¹⁶¹ benutzt.

Aus heutiger Sicht kann das gesamte Programmkonzept als überholt gelten. WordCruncher ist heute so gut wie nicht mehr verfügbar – zumindest nicht in einer Version, die mit aktuellen Betriebssystemen zusammenarbeitet. Die erhobenen Daten sind nicht mehr nutzbar, da WordCruncher ein eigenes Datenformat verwendet, das von keinem anderen Programm unterstützt wird. Viele der von WordCruncher bereit gestellten Funktionen zu deren Realisierung das Programm seinerzeit entwickelt wurde, können heute einfacher mit anderen Programmen ausgeführt werden oder sind, etwa durch die Leistungssteigerung bei Computern, unnötig geworden.

Aus der »Frühzeit« elektronischer Editionen lassen sich weitere Beispiele nennen, beispielsweise die Reihe RECLAMS KLASSIKER AUF CD-ROM, in der von 1996 bis 1998 etwa 30 Titel erschienen sind oder die Reihe LITERAMEDIA des Cornelsen Verlags mit etwa 20 Titeln. In beiden Fällen wurde die Software für Windows 95 entwickelt und läuft nicht mehr zuverlässig auf allen aktuellen Systemen.

Die Frage nach der Alterungsbeständigkeit bzw. der Archivierung und Langzeitbeständigkeit wird oft in Verbindung mit Containermedien wie etwa der CD-ROM aufgeworfen. Dabei werden hauptsächlich zwei Aspekte thematisiert. Zum einen die Beständigkeit des Datenträgers selbst, zum anderen die zukünftige Auslesemöglichkeit solcher Datenträger.

Die Alterungsbeständigkeit optischer Datenträger wie der CD-ROM wird heute auf maximal 50 Jahre geschätzt. Erfahrungswerte liegen dabei noch nicht vor, da die CD-ROM erst seit etwa 20 Jahren zur Verfügung steht. Damit ist sie magnetischen Speichermedien wie Magnetbändern oder Disketten deutlich überlegen. Im Vergleich zu manchen Büchern sind 50 Jahre aber immer noch ein recht kurzer Zeitraum.

Eine andere Frage ist, ob es in 50 Jahren überhaupt noch Geräte geben wird, mit denen man auf CD-ROM/DVD gespeicherte Daten auslesen kann. Dieser zweite Aspekt hat bei der Geschwindigkeit der Technikentwicklung mittlerweile eine größere Bedeutung als der Alterungsbeständigkeit der Datenträger selbst.

2.3.2 Erarbeitung zukunftssicherer elektronischer Editionen

Aus den Erfahrungen mit elektronischen Medien nahe stehenden elektronischen Editionen sollte für zukünftige elektronische Editionen der Aspekt der Nachhaltigkeit bzw. Zukunftssicherheit eine zentrale Rolle spielen. Wie ich dargelegt habe, werden die meisten aktuellen Editionsprojekte, wenn elektronische Medien überhaupt berücksichtigt werden, als buchnahe elektronische Editionen herausgegeben.

160 Vgl. Arbeitsstelle Kant-Ausgabe der Göttinger Akademie der Wissenschaften: http://web.uni-marburg.de/kant//webseite/home_inh.htm (18.05.2008).

161 Musil, Robert: Robert Musil. Der literarische Nachlaß. Hg. von Friedbert Aspetsberger u. a. Elektronisch ediert (Wordcruncher) von Karl Eibl und Marianne Willems. Reinbek bei Hamburg: Rowohlt, 1992.

Dabei mögen wirtschaftliche Überlegungen eine Rolle spielen. Gibt man elektronischen Medien nahe stehende elektronische Editionen auf der Grundlage bestehender Editionen heraus, hat der Verlag meist nur die Kosten für die Retrodigitalisierung und ggf. für die Softwareentwicklung zu tragen. Im Fall von neuen Editionen, die in den meisten Fällen noch als Buchedition konzipiert werden, verursacht die parallele Umsetzung einer eigenständigen elektronischen Edition Mehrkosten. Um diese zu sparen, gibt man elektronischen Publikationsweisen den Vorzug, die mit geringem Aufwand, quasi als Nebenprodukt des Buches, herausgegeben werden können.¹⁶² In den meisten Fällen wird eine Version des gedruckten Buches im *Portable Document Format* als elektronische Version publiziert.

Dokumente im *Portable Document Format* werden meist mit Hilfe des Adobe Readers, einem der zurzeit meistbenutzten Programme, wenn es um die Bereitstellung und Darstellung von Texten in elektronischen Medien geht, benutzt. In erster Linie gewährleistet das *PDF* die originalgetreue Wiedergabe eines Dokuments auf verschiedenen Computersystemen und unterschiedlichen Ausgabemedien. *PDF*-Dokumente können dabei für einzelne Verwendungsarten optimiert werden. Die Möglichkeiten reichen von der Verbreitung über das Internet und die Darstellung am Bildschirm bis zur Druckvorstufe; manche zum Druck eingesetzten Geräte können *PDF* sogar direkt umsetzen. Das *PDF* beruht auf *Postscript*, einer Programmiersprache zur Darstellung von Text und Vektorgrafiken. Im Gegensatz zu *Postscript* ist es jedoch keine Programmiersprache sondern eine *Auszeichnungssprache*.

PDF-Dokumente können von verschiedenen Programmen erzeugt oder angezeigt werden. Der frei für verschiedene Computersysteme verfügbare Adobe Reader dient zum Betrachten und Ausdrucken von *PDF*-Dokumenten. Neben der reinen Darstellung und der Realisierung von *Hypertext*-Funktionen bietet er weitere Möglichkeiten, etwa eine Volltextsuche oder das Anbringen und Kommentieren von Lesezeichen. Somit erfüllt der Adobe Reader einige Grundfunktionen elektronischer Editionen.

Als Datenformat für die Speicherung und Verbreitung formatierter Texte ist das *PDF* zurzeit führend. Durch seine hohe Verbreitung und dadurch, dass Adobe bisher auf eine Abwärtskompatibilität des Datenformats geachtet hat, kann seine Lesbarkeit auf mittlere Sicht als gewährleistet betrachtet werden. Der Nachteil des *PDF* ist, dass es keine semantische Kodierung ermöglicht. Somit lassen sich Texte im *PDF* nur begrenzt mit elektronischen Technologien auswerten.

Eine Alternative wäre es, bei der Arbeit an neuen Editionsprojekten konsequent auf X-Media-Konzepte zu setzen¹⁶³ – angefangen von Erhebung der Arbeitsdaten, über die eigentliche Arbeit an der Edition bis zur Druckvorstufe oder zur elektronischen Edition könnte auf einer gemeinsamen Datengrundlage gearbeitet werden. In diesem Fall stünden alle Arbeitsdaten auch für die Verwendung nach Abschluss der eigentlichen Edition zur Verfügung.

2.3.3 Produktionsseitige Dokumentation

Der dritte Punkt schließlich fand in der wissenschaftlichen Diskussion um elektronische Medien bisher kaum Berücksichtigung. Dabei scheint die Frage, womit sich eine zukünftige Editionswissenschaft beschäftigt, wenn keine Handschriften oder überarbeiteten Typoskripte von Autoren mehr vorliegen, weil diese mit Textverarbeitungsprogrammen gearbeitet haben, durchaus gerechtfertigt. Es wäre Thema einer eigenen Arbeit, zu untersuchen, inwieweit zeitgenössische Autoren die Entwicklung ihrer eigenen Texte dokumentieren, indem sie etwa verschiedene Überarbeitungsstände sichern oder geeignete technische Möglichkeiten nutzen und ob sie sich Gedanken über eine längerfristige Archivierung ihrer Texte und Textentwürfe machen. Einige der im Folgenden angesprochenen Technologien würden sich zumindest für eine solche Dokumentation eignen.

¹⁶² Vgl. 1.3 Buchnahe elektronische Editionen, S. 14.

¹⁶³ Vgl. 2.2.6 X-Media-Publishing, S. 64.

3 Konzepte und Technologien für elektronische Editionen

Ausgangspunkt für diese Arbeit war die Beobachtung, dass es auf dem Gebiet elektronischer Editionen zurzeit zwei unterschiedliche Konzepte gibt: Buchnahe elektronische Editionen, die sich in der Art der Textpräsentation stark am Buch orientieren und meist Einschränkungen bezüglich ihrer Nutzbarkeit in elektronischen Medien bzw. der mit elektronischen Medien möglichen Funktionen in Kauf nehmen, und elektronischen Medien nahe stehende elektronische Editionen, die teilweise umfangreiche elektronische Funktionen zur Verfügung stellen, die aber vom editorischen bzw. wissenschaftlichen Niveau meist nicht an die buchnahen elektronischen Editionen bzw. an aufwändige Bucheditionen heranreichen.

Eine mögliche Erklärung für diese Beobachtung habe ich in der Tradition, Texte auf Papier zu überliefern und zu kommunizieren, vorgestellt, die bis heute den wissenschaftlichen Diskurs auf dem Gebiet der Geisteswissenschaften prägt. Ich habe an einigen Beispielen exemplarisch aufgezeigt, dass Bücher als Informationsträger mehr Informationen vermitteln als der reine Buchstabentext. Informationen sind in Büchern anders »kodierte« als durch die rein »phonetische Qualität« der Buchstaben, aus denen sich die Bedeutung gesprochener Sprache ergibt.¹⁶⁴ Eine wichtige Feststellung dabei war, dass diese Informationen selten expliziert werden und die wissenschaftliche Auseinandersetzung mit diesen Aspekten von Schriftlichkeit eher rudimentär stattfindet.

Buchnahe elektronische Editionen bleiben in vielen Fällen dem »papiernen Denken« verhaftet. Sie übernehmen weitgehend die in Büchern verwendeten Mittel zur Informationsdarstellung und ermöglichen somit den gewohnten Umgang mit Texten. Die vom Buch übernommene Textdarstellung ermöglicht es beispielsweise, Informationen mit Hilfe von Typografie und Layout zu vermitteln oder zu strukturieren.¹⁶⁵ Im elektronischen Medium werden somit Konventionen benutzt, die ursprünglich für das Medium Buch entwickelt wurden. Da eine auf dem Bildschirm dargestellte Buchseite vom Nutzer visuell ähnlich wahrgenommen und ausgewertet werden kann wie eine Buchseite, scheint das Konzept zu funktionieren. Möchte man jedoch elektronische Medien für mehr als zu Darstellungszwecken nutzen, sind buchnahe elektronische Editionen eine Sackgasse, da viele der enthaltenen Informationen nur von einem Leser mit entsprechenden Vorkenntnissen ausgewertet werden können, elektronische Auswertungsmethoden jedoch nicht zur Verfügung stehen. Weitere und neue Zugänge zu Texten, wie sie elektronische Medien theoretisch bieten können, bleiben buchnahen elektronischen Editionen unzugänglich. Dies ist bedauerlich, da gerade neuere Editionen, die aktuelle wissenschaftliche Erkenntnisse und neue editorische Konzepte berücksichtigen, als buchnahe elektronische Editionen herausgegeben werden; erstaunlich, weil bei neuen Editionsprojekten in der Arbeitsphase auf allen Ebenen elektronisch gearbeitet wird und erklärlich, da als Richtlinie für die Arbeit immer noch eine Buchedition steht, die den Bedingungen des Mediums Buch unterworfen ist.

Aber auch elektronischen Medien nahe stehende elektronische Editionen schöpfen kaum das Potential elektronischer Datenverarbeitung aus. Selbst wenn hier viele Möglichkeiten, beispielsweise beim Information *Retrieval*, zur Verfügung stehen, bleibt der praktische Nutzen hinter den theoretischen Möglichkeiten zurück, da der Informationsgehalt der Texte geringer ist: Elektronischen Medien nahen elektronischen Editionen liegen meist ältere, etablierte Buchausgaben zugrunde, die nicht unbedingt dem neuesten wissenschaftlichen Stand entsprechen. Daneben beschränkt man sich bei der Retrodigitalisierung vorhandener Ausgaben meist auf den vorliegenden Textbestand. Weitergehende Informationen über die Buchausgabe werden nicht erfasst, das Material bzw. die der Edition zugrundeliegenden Informationen stehen nicht mehr zur Verfügung. Informationen, die Buchausgaben auf anderen Ebenen als der Schriftebene enthalten, werden selten berücksichtigt.

164 Dafür gehen dem Text beim Übergang von der Mündlichkeit zur Schriftlichkeit orale Aspekte, etwa die Betonung, verloren.

165 Vgl. 2.1.3 Text und Gestaltung, S. 43.

Wünschenswert wäre eine Synthese aus beiden Editions Konzepten. Der Informationsgehalt buchnaher elektronischer Informationen müsste elektronischen Mitteln zugänglich gemacht werden. Informationen, die Bücher etwa typografisch vermitteln, müssen mit elektronischen Mitteln kodiert werden. Dazu müssen andere Informationsebenen von Büchern zuerst explizit dargestellt und anschließend für elektronische Medien kodiert werden. In historisch-kritischen Bucheditionen beispielsweise werden Befunde meist typografisch kodiert oder explizit im Apparat dargestellt. Im elektronischen Medium wäre es naheliegend, diese Informationen nicht zuerst in einen typografischen Kode zu übersetzen und das Ergebnis anschließend in einem Datenformat zu kodieren, das wiederum die typografischen Codes elektronisch speichert, sondern nach dem Prinzip der semantischen Kodierung zu verfahren und die entsprechenden Informationen unmittelbar mit dem Text zu kodieren. In einem geeigneten elektronischen Datenformat wären diese Informationen offen für aktuelle elektronische Zugriffsmöglichkeiten, im Idealfall auch für zukünftige Technologien, die sich zurzeit bereits abzuzeichnen beginnen. Auf der anderen Seite stellt sich die Frage, ob es notwendig ist, im elektronischen Medium an Konventionen des Buchdrucks festzuhalten. Die Art der Darstellung, also die Form, in der die kodierten Informationen später auf dem Bildschirm dargestellt werden, muss nicht wie beim Buchdruck vorgegeben werden, sondern kann durch den Nutzer selber beeinflusst werden. Diese Möglichkeit bedeutet in einem Punkt ein Umdenken bei der editorischen Arbeit.

Im Rahmen einer Buchausgabe gehört es zu den Aufgaben des Herausgebers, ein Konzept zu entwickeln, wie die Textbefunde mit den Mitteln der Textgestaltung in Büchern dargestellt werden können. An diese Stelle tritt bei einer elektronischen Edition die Entwicklung eines Konzepts zur Textkodierung, das es ermöglicht, die Informationen über die Befunde zu kodieren und Anknüpfungsmöglichkeiten für die Darstellung, Verarbeitung und Auswertung in elektronischen Medien bietet. Bei konsequenter Umsetzung können damit nahezu alle Informationen, die ein Buch enthält, im elektronischen Medium zur Verfügung gestellt werden. Dabei muss allerdings der Schritt von der Kodierung mit Hilfe von Schrifttechnologien zu einer Kodierung mit Hilfe elektronischer Kodierungen gemacht werden. Damit dabei keine Informationen verloren gehen, müssen alle Informationsebenen bei der Umkodierung berücksichtigt werden. Welche Schwierigkeiten sich dabei auftun, habe ich bereits dargelegt.

Die Aufmerksamkeit, die bei der editorischen Arbeit an einer Buchausgabe der Gestaltung (Typografie, Layout, Design) oder anders ausgedrückt, dem Druckformat zuteil wird, muss im Falle einer elektronischen Edition dem Datenformat zukommen. Dazu kommt, dass Informationen, die ein Buch implizit enthält (etwa Verfügbarkeit, Aussehen und Zustand eines Buches), in elektronischen Medien explizit angegeben werden müssen, da sie mit dem Medienwechsel verloren gehen oder sich im elektronischen Medium anders darstellen.

Im Hinblick auf Bücher muss darüber hinaus ein weiterer wichtiger Aspekt bei der Kodierung von Texten berücksichtigt werden: Die Nachhaltigkeit des Datenformats. Unter Berücksichtigung der schnellen Entwicklung auf dem Gebiet der elektronischen Datenverarbeitung sollten kodierte Daten eine lange »Haltbarkeit« aufweisen. Dies ist am einfachsten zu gewährleisten, indem die kodierten Texte mit möglichst wenigen technischen Voraussetzungen lesbar bleiben.

Damit elektronische Editionen den Ansprüchen, die an Bucheditionen gestellt werden, genügen können, müssen verschiedene Punkte berücksichtigt werden:

- dauerhafte Speicherung des Textes an sich
- Möglichkeiten, auch komplexe Informationen zum Text zu kodieren
- flexible, bedarfsgerechte Darstellungsmöglichkeiten
- Schnittstellen zur Auswertung mit elektronischen Mitteln und Methoden

In vielen Bereichen elektronischer Informationsverarbeitung haben sich *Auszeichnungssprachen* zur Kodierung elektronischer Texte durchgesetzt. In Verbindung mit der Erhebung von Metadaten, einer Praxis, die auch im Kontext von Büchern, etwa bei der Katalogisierung, gebraucht wird, lassen sich Konzepte entwickeln, die den aufgeführten Anforderungen entsprechen.

Im Folgenden stelle ich zunächst die Prinzipien *Auszeichnungssprachen* und Metadaten vor. Anschließend gehe ich auf die zurzeit meistbeachtete *Auszeichnungssprache*, die *Extensible Markup Language* ein. Dabei werde ich zunächst allgemeine Grundlagen erläutern. Anschließend werde ich als eine Anwendung von XML die Richtlinien der *Text Encoding Initiative (TEI)* zur Kodierung von Texten in den Geisteswissenschaften vorstellen und damit verbundene Möglichkeiten und Probleme anhand zweier Beispiele aufzeigen. Anschließend werde ich auf Metadaten und geeignete Kodierungsmöglichkeiten eingehen.

3.1 Auszeichnungssprachen

Um den problemlosen Austausch von Daten zwischen den unterschiedlichen Computersystemen in einer vernetzten Welt sowie zwischen verschiedenen Anwendungsprogrammen zu ermöglichen, wurde und wird weltweit an einheitlichen Datenformaten gearbeitet. Im Bereich von Texten haben sich dabei standardisierte *Auszeichnungssprachen*, englisch Markup Languages (ML) durchgesetzt.

3.1.1 Grundprinzip von Auszeichnungssprachen

Der Grundgedanke der meisten heute gebräuchlichen *Auszeichnungssprachen* besteht darin, mit einer begrenzten Menge an Zeichen komplexe Informationsstrukturen abzubilden. Hintergrund ist, dass man in den Anfängen der Computertechnik nur eine begrenzte Anzahl von Zeichen zur Verfügung hatte, da alle Computer auf der Grundlage der sogenannten Digitaltechnik arbeiten.

Grundlagen der Digitaltechnik

Die Digitaltechnik beruht auf dem Konzept, Informationen als Abfolge zweier Zustände abzubilden: »0« (Null) und »1« (Eins). Diese zwei Zustände bilden die kleinste digitale Informationseinheit, das Bit. In einem elektrischen Leiter lassen sich »1« und »0« einfach durch die Zustände »Strom« bzw. »kein Strom« darstellen, auf einer Lochkarte etwa »Loch« und »kein Loch«.

Um komplexere Informationen darzustellen, bildet man aus mehreren Bits sogenannte Bitwörter. Die Anzahl der damit darstellbaren Informationen wächst dabei mit jedem Bit innerhalb eines Wortes um die Potenz der Anzahl der Bits zur Basis Zwei. Mathematisch gesehen heißt das: Wenn n die Anzahl der Bits in einem Wort ist, so ist die Anzahl der darstellbaren Zustände 2^n .

Bits	Zustände	Potenz zu 2	Anzahl der Zustände
1	0 1	2^1	2
2	00 01 10 11	2^2	4
3	000 001 010 011 100 101 110 111	2^3	8
4	...	2^4	16
8		2^8	256
16		2^{16}	65536

Tabelle 1. Darstellung der Zustände in Abhängigkeit von der Anzahl der Bits

Eine Informationseinheit aus 8 Bits bzw. ein Acht-Bit-Wort wird Byte genannt und ist die heute übliche Größe digitaler Informationseinheiten. Mit einem Byte können 2^8 , also 256 Zustände beschrieben werden.

Auf der Ebene digitaler Informationsverarbeitung werden Bitwörter dadurch realisiert, dass eine entsprechende Anzahl von Bits parallel verarbeitet wird, etwa durch eine entsprechende Anzahl von Leitungen. Um ein Byte elektronisch zu verarbeiten, benötigt man demnach acht Leitungen. Zur Datenverarbeitung wurden früher meist nur 7 Bit verwendet, womit 128 Zustände verarbeitet werden konnten. Auf der Basis dieser 128 verfügbaren Zustände wurde der *American Standard Code for Information Interchange (ASCII)* entwickelt, der heute die Grundlage der Zeichenkodierung für beinahe alle digitalen Systeme ist. Der *ASCII*-Kode sieht 33 so genannte Steuerzeichen und 95 darstellbare Zeichen vor: Die Buchstaben von A bis Z in Groß- und Kleinschreibung, Ziffern und einige Sonderzeichen. Deutsche Umlaute werden vom 7 Bit-*ASCII*-Kode nicht berücksichtigt. Auch wenn es mittlerweile umfangreichere Zeichenkodes gibt, die auf 16 Bits beruhen – anzuführen ist hier vor allem der *UNICODE*, der sich weltweit als Standard etabliert und zurzeit mehr als 99000 Zeichen erfasst –, wurden die grundlegenden Technologien von *Auszeichnungssprachen* entwickelt, als man auf 7 Bit *ASCII* beschränkt war.

Grundlagen der Textauszeichnung

Die Auszeichnung von Texten ist keine Erfindung des Computerzeitalters. Bereits in der Antike wurden Texte mit Markierungen für den mündlichen Vortrag ausgezeichnet.¹⁶⁶ Ein System von Auszeichnungen wurde für die Anweisungen zur Einrichtung des Drucksatzes im dem Druck zugrundeliegenden Manuskript entwickelt. Dazu wurde der auszuzeichnende Bereich meist unterstrichen und am Rand des Textes vermerkt, wie der markierte Text gedruckt werden soll, etwa kursiv, fett etc. Diese Form der Auszeichnung setzt voraus, dass der auszuzeichnende Text auf ein Trägermaterial – Papier – geschrieben ist, das in die Textauszeichnung einbezogen wird: Unterhalb der Schrift muss Platz für Markierungen bzw. Untersteichungen sein, am Rand Raum für Vermerke.

Ein Beispiel für die Textauszeichnung auf Papier sind etwa die Korrekturzeichen, für die es in Deutschland die Norm DIN 16511 gibt. Als Beispiel für die Möglichkeiten der Textauszeichnung soll im Folgenden eine Zeile aus Büchners Drama *DANTONS TOD* dienen:

Danton. Spazieren, mein Junge, spazieren! *(er geht)*

Die dargestellte Zeile ist bereits typografisch kodiert. Der Name »Danton« ist fett gedruckt. Im Kontext mit dem darauf folgenden Text, der kein Satz, aber eine sprachliche Aussage ist, nimmt jeder typografisch gebildete Leser sofort an, dass es sich um eine Zeile aus einem Drama handeln muss, wobei Danton der Sprecher ist. Das anschließende, kursiv in Klammern angefügte »er geht« wird dabei als Regieanweisung interpretiert.

Würde die angeführte Zeile in einem Typoskript, für das nur eine Schriftart zur Verfügung stand, erscheinen, könnte eine Textauszeichnung, die angibt, wie die Seite später gedruckt aussehen soll, so aussehen:

Danton. Spazieren, mein Junge,
spazieren. (er geht)

fett
kursiv

Beispiel 1. Textauszeichnung in einem Typoskript

Textauszeichnung in elektronischen Medien

Texte in elektronischen Medien liegen nicht auf einem Trägermaterial vor, sondern existieren in Form elektronischer Informationseinheiten. Außerhalb der technischen Möglichkeiten gibt es keinen »Raum« für Auszeichnu

¹⁶⁶ Vgl. Entwicklung typografischer Gliederungsmerkmale, S. 44.

ngsmarkierungen bzw. Anmerkungen, wie es etwa das Papier im Falle geschriebener Texte bietet. Zur Kodierung von Auszeichnungen steht eine begrenzte Anzahl von Zeichen zur Verfügung, die durch den Zeichenkode, in den meisten Fällen der *ASCII*-Kode, vorgegeben wird.

Um unter diesen Bedingungen die Möglichkeit zu schaffen, zusätzliche Informationen über einen Text zu kodieren, werden in elektronischen Medien häufig bestimmte Zeichen als Markierungszeichen verwendet. Das Auftauchen eines bestimmten Zeichens, das als Markierung definiert ist, besagt: Hier steht jetzt etwas, das nicht zum eigentlichen Text gehört.

Neben dem Beginn einer Auszeichnungsmarkierung muss auch ihr Ende gekennzeichnet werden. Dies kann einfach durch ein Leerzeichen oder einen Zeilenumbruch¹⁶⁷ geschehen. Oft werden zu diesem Zweck aber korrespondierende Zeichen verwendet, da hier der Zusammenhang, der Anfang und das Ende der Auszeichnungsmarkierung, besser zu erkennen sind. Daneben ist es vorteilhaft, wenn diese Zeichen in den auszuzeichnenden Texten selten verwendet werden. Wenn ein Zeichen zur Kennzeichnung einer Auszeichnungsmarkierung verwendet wird, darf es im sonstigen Text nicht in seiner »normalen« Bedeutung verwendet werden. Die Auszeichnung wäre sonst nicht eindeutig, da sich bei jedem Vorkommen des Zeichens die Frage stellen würde, ob das Zeichen nun in seiner ursprünglichen Bedeutung oder als Beginn einer Auszeichnungsmarkierung verwendet wird. Aus diesem Grund werden gern geschweifte (»{ }«) oder eckige (»[]«) Klammern verwendet, die sonst selten benutzt werden. Im Falle der Gruppe von *Auszeichnungssprachen*, auf die ich im Folgenden eingehen werde, hat man sich für die Zeichen »kleiner als« und »größer als« entschieden, die im Sinne von spitzen Klammern (»< >«) benutzt werden.

Auszeichnungssprachen und semantische Kodierung

Die Möglichkeiten der Textauszeichnung in elektronischen Medien können auf verschiedene Arten genutzt werden. Zur Textdarstellung wurden Datenformate entwickelt, bei denen der Schwerpunkt der Textauszeichnung auf der Formatierung liegt. Sollen Informationen jedoch nicht nur angezeigt, sondern auch mit elektronischen Mitteln ausgewertet werden, bietet sich eine semantische Kodierung an.¹⁶⁸ Zur Veranschaulichung ziehe ich ein Beispiel heran. In dem Satz

Der Baum ist **belaubt**.

habe ich das Wort »belaubt« rot geschrieben. Im schwarzweißen Druck ist die Farbe nicht zu erkennen. Wenn im Druck keine Farben zur Verfügung stehen, muss die Information, dass das Wort »belaubt« rot dargestellt sein soll, anders übermittelt werden. Explizit in den Text zu schreiben, dass der Text rot sein soll, verbietet sich, da der Satz sonst eine andere Bedeutung bekommt:

Der Baum ist rot belaubt.

Hier ist nicht erkennbar, dass sich die Angabe »rot« auf die Darstellung des Wortes »belaubt« und nicht auf die Farbe des Laubes bezieht. Die Information »rot« bezieht sich auf eine andere Ebene des Textes, die der Textdarstellung. Um diese Informationsebene der Textdarstellung übermitteln zu können, ohne dass sie sich mit den Informationen der Textebene vermischt, kann man die Information kodieren. Dazu muss explizit ein Kode vereinbart werden. Eine solche Vereinbarung könnte lauten: »Eine Unterstreichung im Text zeigt an, dass das Wort rot geschrieben ist«. Damit wäre der Kode bekannt und der Satz könnte folgendermaßen dargestellt werden:

Der Baum ist belaubt.

Diese Art der Kodierung habe ich weiter oben als »typografische Kodierung« bezeichnet, da sie typografische Mittel benutzt, um Informationen zu vermitteln, die nicht unmittelbar dargestellt werden können.

167 Auch zur Kodierung von Zeilenumbrüchen gibt es einen ASCII-Kode, der zu den sogenannten Steuerzeichen gehört.

168 Vgl. 3.1 Auszeichnungssprachen, S. 71 und 2.2.5 Semantische Kodierung, S. 63.

Stehen zur Kodierung auch keine typografischen Mittel zur Verfügung, wie es etwa der Fall ist, wenn der Satz nur mit Hilfe von Zeichendaten übermittelt werden soll, muss die Information anders im Text hinterlegt werden. Hier bietet sich eine Kodierung auf der Grundlage von Ausnahmezeichen an. Damit die Information richtig ausgewertet werden kann, müssen auch hier Vereinbarungen getroffen werden, etwa: »Spitze Klammern stellen Ausnahmezeichen dar. Informationen, die zwischen spitzen Klammern stehen, beziehen sich auf die Textdarstellung«. Damit würde der Satz folgendermaßen aussehen:

Der Baum ist <rot>belaubt</rot>.

Für das letzte Beispiel habe ich vorgegriffen und mich an der Textauszeichnung gemäß der *Extensible Markup Language* orientiert.¹⁶⁹ Neben dem Anfang muss auch das Ende des Bereiches, der rot dargestellt werden soll, gekennzeichnet werden. Gemäß *XML* kennzeichnen »<« und »>« Beginn und Ende einer einzelnen Auszeichnungsmarkierung, die gleiche Markierung, bei der nach der öffnenden spitzen Klammer ein Schrägstrich folgt, zeigt das Ende des markierten Bereichs an.

Der Vorteil dieser Form der Auszeichnung ist, dass sich durch das Vereinbaren weiterer Konventionen auch andere Informationsebenen kodieren lassen. So kann man etwa angeben, auf welchen Aspekt sich die Information in der Auszeichnung bezieht. Ich greife vor und stelle ein Modell vor, indem sowohl die Textinformationen und die Informationen über die Darstellung als auch Informationen über die Grammatik des Satzes kodiert sind. Dazu lege ich fest, dass innerhalb der Auszeichnung zuerst die Informationsebene genannt wird, auf die sich die Angabe bezieht, danach den Aspekt und die entsprechende Information. Das Beispiel richtet sich bereits nach den in *XML* üblichen Konventionen für *Attribute*:

```
<Grammatik Wortart="Artikel">Der</Grammatik>
<Grammatik Wortart="Substantiv">Baum</Grammatik>
<Grammatik Wortart="Verb">ist</Grammatik>
<Grammatik="Adjektiv">
  <Darstellung Schriftfarbe="rot">
    belaubt
  </Darstellung>
</Grammatik>.
```

Das Beispiel ist schon recht komplex. Auf die Regeln der *XML*-Kodierung gehe ich noch detailliert ein. Das obige Beispiel verdeutlicht, wie man Informationen, die verschiedene Ebenen eines Textes betreffen, mit Hilfe eines Prinzips kodieren kann.

Deutlich wird an dieser Stelle auch, wie unsinnig es ist, in elektronischen Medien an den Konventionen der Druckdarstellung festzuhalten. Die betreffende Textstelle für elektronische Medien

Der Baum ist <unterstrichen>belaubt</unterstrichen>.

auszuzeichnen und durch einen Vermerk zu ergänzen, dass Wörter, die als unterstrichen markiert sind, rot dargestellt werden, hieße, die typografisch kodierte Information wiederum elektronisch zu kodieren. Grundlage der Kodierung in elektronischen Medien sollte immer eine semantische Kodierung sein.¹⁷⁰

3.1.2 Standard Generalized Markup Language

Einer der am weitesten verbreiteten Standards zur Kodierung von Texten und anderen Informationen ist die *Standard Generalized Markup Language (SGML)*, die 1986 als ISO Standard 8879 zertifiziert wurde. *SGML* basiert auf dem bereits 1969 bei IBM entwickelten Konzept einer *General Markup Language (GML)*.¹⁷¹ Bereits bei der Entwicklung von *SGML* stand die Auszeichnung von Texten im Vordergrund:

¹⁶⁹ Vgl. 3.2.1 Einführung in *XML*, S. 81.

¹⁷⁰ Vgl. 2.2.5 Semantische Kodierung, S. 63.

¹⁷¹ *GML* hat sich jedoch nicht durchgesetzt. Heute wird das Kürzel *GML* für die *Geography Markup Language* benutzt – einer *XML*-basierten Auszeichnungssprache zur Kodierung geografischer Daten.

SGML can be used for publishing in its broadest definition, ranging from single medium conventional publishing to multi-media data base publishing. SGML can also be used in office document processing when the benefits of human readability and interchange with publishing systems are required.¹⁷²

SGML selber ist keine fest definierte *Auszeichnungssprache* im eigentlichen Sinn, sondern eine Metasprache, eine Grammatik, auf deren Grundlage konkrete *Auszeichnungssprachen* entwickelt werden können. SGML stellt die Grammatik und in Verbindung mit der Definition eines Vokabulars entsteht eine *Auszeichnungssprache*. Konkrete *Auszeichnungssprachen* bezeichnet man auch als Anwendungen von SGML.

SGML bzw. entsprechende SGML-Anwendungen kommen seit geraumer Zeit bei der Herstellung von Printmedien zum Einsatz, etwa in Redaktionssystemen von Zeitungen und anderen, textorientierten Datenbankanwendungen (wie z. B. technischen Dokumentationen, Telefonbüchern etc.) und bei der Übernahme entsprechender Daten in die Druckvorstufe. Die Entwicklung geeigneter Anwendungen bzw. der Auszeichnungen auf Grundlage von SGML fällt dabei meist in den Bereich der Technik bzw. Informatik, der von geisteswissenschaftlicher Seite gern als »Black Box« betrachtet wird. Obwohl hier im Hinblick auf die Textauszeichnung schon eine weitreichende Entwicklung stattgefunden hat, wurde diese vom philologischen Standpunkt aus wenig beachtet bzw. wird für nicht beachtenswert gehalten.

Im Rahmen elektronischer Editionen wird SGML etwa bei den Textdatenbanken der Firma Chadwyck-Healey eingesetzt.¹⁷³ Auf der Grundlage von SGML wurde ein Datenformat entwickelt, in dem die Textdaten für die Übernahme in die Textdatenbank kodiert werden. Auf die SGML-kodierten Texte selber hat ein Nutzer keinen Zugriff. Er sieht lediglich die Darstellung, die mit Hilfe der Benutzerseite der Software, des sogenannten »Frontend«, generiert wird. Der in Abbildung 4 (S. 18) Als Screenshot abgebildeter Text der elektronischen Kafka-Ausgabe sieht im zugrunde liegenden SGML-kodierten Text folgendermaßen aus:

```
<pb n=11&idpage=KKAPp11&pn=11&><div name="Der Onkel / Leni" somhead="Der
Proce&B: Der Onkel / Leni" somlevel=2 comhead="Der Proce&B: Der Onkel / Leni">
<head pn=11&>Der Onkel <1b>Leni</head>
<xref RID=007EE11& type=ALT text="Editorische Eingriffe" pn=11&>
<xref RID=007VA11& type=ALT text="Varianten" pn=11&>
<p pn=11&><pp pn=11&><ln n=3>Eines Nachmittags&mdash;K. war gerade vor dem
Postabschlu&esset;
<ln n=4>sehr besch&aulm;ftigt&mdash;dr&aulm;ngte sich zwischen zwei
<ln n=5>Dienern, die Schriftst&uuml;cke hereintrugen K.'s Onkel Karl,
<ln n=6>ein kleiner Grundbesitzer vom Lande, ins Zimmer. K.
<ln n=7>erschrak bei dem Anblick weniger, als er schon vor
<ln n=8>1&aulm;ngerer Zeit bei der Vorstellung vom Kommen des
<ln n=9>Onkels erschrocken war. Der Onkel mu&esset;ite kommen,
<ln n=10>das stand bei K. schon etwa einen Monat lang fest.
<ln n=11>Schon damals hatte er ihn zu sehen geglaubt, wie er ein
<ln n=12>wenig geb&uuml;ckt, den eingedr&uuml;ckten Panamahut in der
<ln n=13>Linken die Rechte schon von weitem ihm entgegenstreckte
<ln n=14>und sie mit r&uuml;cksichtsloser Eile &uuml;ber den
<ln n=15>Schreibtisch hin reichte, alles umsto&esset;end, was ihm im
<ln n=16>Wege war. Der Onkel befand sich immer in Eile, denn er
<ln n=17>war von dem ungl&uuml;cklichen Gedanken verfolgt, bei
<ln n=18>seinem immer nur eint&aulm;gigen Aufenthalt in der Hauptstadt
<ln n=19>m&uuml;sse er alles erledigen k&ouml;nnen, was er sich vorgenommen
<ln n=20>hatte und d&uuml;rfe &uuml;berdies auch kein gelegentlich
```

172 ISO 8879:1986, Clause 0 »This International Standard specifies a language for document representation referred to as the „Standard Generalized Markup Language“ (SGML). SGML can be used for publishing in its broadest definition, ranging from single medium conventional publishing to multi-media data base publishing. SGML can also be used in office document processing when the benefits of human readability and interchange with publishing systems are required.«

173 Chadwyck Healey ist in Deutschland vornehmlich durch die Firma ProQuest vertreten. In Zusammenarbeit mit etablierten Verlagen werden in Deutschland digitale Werkausgaben wie Kafka, Goethe, Schiller oder Thomas Mann auf CD-ROM oder im WWW vertrieben.

```

<ln n=21>sieh darbietendes Gespr&auml;ch oder Gesch&auml;ft oder
  Vergn&uuml;gen
<ln n=22>sieh entgegen lassen. Dabei mu&eset;ite ihm K., der ihm
<ln n=23>als seinem gewesenen Vormund besonders verpflichtet</pp>
<pblink pn=119><pb n=119 idpage=KKAPP119 pn=119>
<xref RID=007VA119 type=ALT text="Varianten"></pblink><pp pn=119>
<ln n=1>war, in allem m&ouml;glichen behilflich sein und ihn au&eset;ierdem
<ln n=2>bei sich &uuml;bernachten lassen. &ldquo;Das Gespenst vom
<ln n=3>Lande&rdquo; pflegte er ihn zu nennen.</pp></p>
<p pn=119><pp pn=119>&indent;<ln n=4>Gleich nach der Begr&uuml;e&eset;ung&mdas
hisich in das Fauteuil zu
<ln n=5>setzen, wozu ihn K. einlud, hatte er keine Zeit&mdash;bat

```

Beispiel 2. Auszug aus dem SGML-Quellcode der elektronischen Editionen Kafkas Werke in WWW und Kafkas Werke auf CD-ROM¹⁷⁴

Auch ohne *SGML*-Kenntnisse ist der Text weitgehend lesbar, bei näherer Betrachtung erschließen sich Einzelheiten der Kodierung. So ist die von der Buchvorlage übernommene Seiten- und Zeilenzählung gut zu erkennen.

3.1.3 Hyper Text Markup Language

Eine der bekanntesten *SGML*-Anwendungen ist die *Hyper Text Markup Language* (*HTML*). *HTML* wurde ursprünglich zum Datenaustausch im *World Wide Web* entwickelt und vom World Wide Web Consortium (W3C),¹⁷⁵ einem Gremium, das zur Standardisierung der das *World Wide Web* betreffenden Techniken etabliert wurde, gepflegt. Die letzte vom W3C verabschiedete Version war *HTML* 4.01. Die Weiterentwicklung von *HTML* erfolgt nicht mehr auf der Basis von *SGML*, sondern von *XML* (s. u.) und wird als *Extensible Hyper Text Markup Language* (erweiterbares *HTML*, *XHTML*) bezeichnet.

Das Konzept von *HTML* besteht aus einer Mischung von formaler und semantischer Textauszeichnung. Einige Auszeichnungsmarkierungen, beispielsweise für Überschriften verschiedener Gliederungsebenen, zur Markierung von Zitaten und Quellen oder zur Kennzeichnung von Definitionen, dienen sowohl der inhaltlichen Kategorisierung des entsprechend ausgezeichneten Textes, als auch der Formatierung des Textes. Daneben gibt es Markierungen, die ausschließlich der Formatierung von Text dienen.

Das oben angeführte Beispiel aus DANTONS TOD würde *HTML*-kodiert etwa folgendermaßen aussehen:

```

<p>
  <b>Danton.</b>
  Spazieren, mein Junge, spazieren!
  <i>(er geht)</i>
</p>

```

Wird der Text um das entsprechende *HTML*-Rahmengerüst ergänzt, wird er in nahezu jedem Browser so oder ähnlich dargestellt:

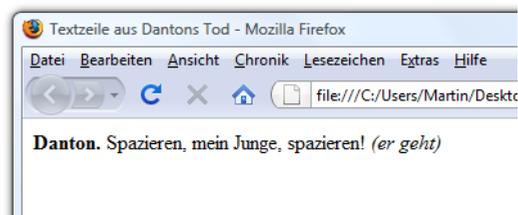


Abbildung 20. Beispiel für die Anzeige der *HTML*-kodierten Textzeile aus DANTONS Tod im Internet-Explorer 8

174 Auszug aus dem *SGML*-kodierten Text der Ausgabe Kafkas Werke im WWW / Kafkas Werke auf CD-ROM. © 1999 Chadwyck-Healey Ltd, Cambridge.

175 <http://www.w3.org/> (18.05.2008).

Das Prinzip der *HTML*-Auszeichnung ist einfach zu verstehen: »p« kennzeichnet eine Textzeile (paragraph), »b« Text, der fett darzustellen ist (bold), »i« Text, der kursiv dargestellt werden soll (italic).

Elektronische Editionen, die auf *HTML* basieren, gibt es wenige. Das PROJEKT GUTENBERG¹⁷⁶ publiziert *HTML* kodierte Texte im Internet. Für anspruchsvolle elektronische Editionsprojekte ist *HTML* nicht gut geeignet, da die Kodierungsmöglichkeiten sehr beschränkt sind.¹⁷⁷

Die Stärke von *HTML* ist, dass sie *Hypertext*-Technologien unterstützt. Daneben erlaubt *HTML* es, andere Daten, etwa Bilder oder Ton recht einfach in *HTML*-Dokumente zu integrieren und unterstützt damit multimediale Anwendungen.

Das ursprüngliche Konzept von *HTML* entpuppte sich schnell als unzureichend. Heute stehen bei der Verwendung von *HTML* gestalterische Aspekte im Vordergrund.

3.1.4 Extensible Markup Language

Die *Extensible Markup Language (XML)* wurde ursprünglich für Internet-Anwendungen entwickelt, bei denen eine *SGML*-konforme Auszeichnung zu komplex erschien. Wie *SGML* ist *XML* eine Metasprache, auf deren Grundlage konkrete Anwendungen entwickelt werden können. Gepflegt wird der Standard vom World Wide Web Consortium.

In vielen Bereichen hat *XML* inzwischen *SGML* abgelöst. Zum einen kann *XML* flexibler eingesetzt werden, da sie verschiedene Stufen an Komplexität zulässt und der Aufwand bei der *XML*-Kodierung bedarfsgerecht auf verschiedenen Niveaus erfolgen kann.¹⁷⁸ Zum anderen, weil *XML* durch weitere Technologien, die auf die *XML*-Strukturen zurückgreifen oder auf ihnen aufbauen, ergänzt oder erweitert werden kann. Daneben unterstützt *XML* im Gegensatz zu *SGML* die Verwendung aktueller Zeichenkodierungen wie etwa *UNICODE*.

Im Hinblick auf Texte sehen *XML*-basierte Konzepte die Trennung von Form und Inhalt vor. Dabei steht meist eine inhaltliche Auszeichnung im Vordergrund der *XML*-Kodierung, da die Darstellung der kodierten Texte durch eine weitere Instanz, meist ein sogenanntes *Stylesheet*, gesteuert wird, etwa *Cascading Stylesheets (CSS)* oder eine Ausgabesteuerung auf Grundlage der *Extensible Stylesheet Language (XSL)*.¹⁷⁹

Geisteswissenschaftliche elektronische Editionen, die völlig auf *XML* und *XML*-Technologien beruhen, gibt es derzeit im deutschsprachigen Raum nicht. Wie *SGML* stellt *XML* aber in einigen Fällen das Datenformat dar, auf dem elektronische Editionen aufbauen. So beruhen beispielsweise die Editionen der DIGITALEN BIBLIOTHEK des Verlags Directmedia auf *XML*-kodierten Texten, die, ähnlich wie die elektronischen Editionen von Chadwyck-Healey, durch eine spezielle Software für den Nutzer aufbereitet werden. Die elektronische Edition DER JUNGE GOETHE IN SEINER ZEIT¹⁸⁰ enthält die Texte nicht nur im proprietären Datenformat der verwendeten Folio Views Software, sondern auch in einer nach den Richtlinien der *Text Encoding Initiative*¹⁸¹ *XML*-kodierten Version.

Auf der Grundlage von *XML* lassen sich beliebige Arten der Textauszeichnung entwerfen. Ohne dass eine spezielle *XML*-Anwendung, also eine ausformulierte *Auszeichnungssprache* mit *XML*-Grammatik zugrunde gelegt wird, kann das angeführte Beispiel nach Belieben kodiert werden. Denkbar ist etwa eine formale Auszeichnung:

176 <http://gutenberg.spiegel.de/> (18.05.2008).

177 Beispiele für *HTML*-kodierte Texte findet man im Projekt Gutenberg unter der o. g. URL. Im Browser kann man sich den Quelltext jedes Werkes anzeigen lassen.

178 Ein *XML*-Dokument kann »wohlgeformt« sein, d. h. es erfüllt die grundlegendsten Formalien der *XML*-Auszeichnung; oder »gültig« sein, dann entspricht es einem konkret festgelegten Inhaltsmodell, etwa einer DTD, einem *XML*-Schema oder einem RELAX NG Schema. Vgl. 3.2.4 *XML*-Inhaltsmodelle, S. 87.

179 Vgl. Technologien zur Formatierung, S. 92.

180 Goethe, Der junge Goethe in seiner Zeit.

181 Vgl. 3.3 Die Text Encoding Initiative, S. 94.

```
<Zeile>
  <fett>Danton.</fett>
  Spazieren, mein Junge, spazieren!
  <kursiv>(er geht)</kursiv>
</Zeile>
```

Aber auch eine semantische Auszeichnung ist denkbar:

```
<Sprechtext>
  <Sprecher>Danton.</Sprecher>
  Spazieren, mein Junge, spazieren!
  <Regieanweisung>(er geht)</Regieanweisung>
</Sprechtext>
```

Während sich die formale *XML*-Auszeichnung im Prinzip kaum von der *HTML*-Auszeichnung unterscheidet – die Auszeichnungsmarkierungen beschreiben nur die Formatierung des Textes –, hat die semantische Textauszeichnung den Vorteil, dass die Funktionen der Textteile klar werden. Fett könnte im Drama auch eine Überschrift gesetzt sein. Die Kodierung als »Sprecher« markiert dagegen eindeutig die Funktion innerhalb des Dramas.

3.1.5 XML-basierte Datenformate

Beim technischen Umgang mit Texten haben sich *Auszeichnungssprachen* auf breiter Ebene durchgesetzt. Besonders im Zusammenhang mit dem sogenannten X-Media-Publishing,¹⁸² der Publikation gleicher Inhalte in verschiedenen Formen und in verschiedenen Medien (Internet, Datenbanken, CD-ROM, eBook, Druck etc.), dienen *Auszeichnungssprachen* als universelles Speicher- und Austauschformat. Um nicht für jedes Medium oder jede Verwendung einen eigenen, an die jeweiligen technischen Voraussetzungen angepassten Text vorhalten zu müssen, was bei Veränderungen des zugrundeliegenden Textes beispielsweise dazu führt, dass diese Veränderungen jeweils an allen Ausgangstexten gleichermaßen vorgenommen werden müssen, greift man beim Konzept des X-Media-Publishing auf eine einheitliche, medienneutrale Textgrundlage zurück.

Verschiedene Softwarehersteller haben für ihre Programme *XML*-Schnittstellen entwickelt. Neben Satzprogrammen wie QuarkXPress oder InDesign, die seit geraumer Zeit Möglichkeiten zum Im- und Export von *XML*-Daten bieten, wird *XML* auch von Programmen für Büroanwendungen, sogenannten Office-Programmen, unterstützt. Das mit Word 2007 neu eingeführte »docx«-Speicherformat für Word Dokumente basiert ebenso auf *XML* wie das Open XML Format, mit dem die Firma Microsoft gerade versucht, ein neues, universelles Austauschformat für Office-Programme zu etablieren.¹⁸³ Das gleiche Ziel verfolgt die »Organization for the Advancement of Structured Information Standards« (OASIS), eine internationale, nicht-gewinnorientierte Organisation, die sich mit der Weiterentwicklung von E-Business- und Web-Service-Standards beschäftigt, mit dem »OpenDocument«-Format. Das »OpenDocument« Speicherformat wird bisher vornehmlich von der auf Unix-/Linux-Plattformen verbreiteten Open Office Software unterstützt.¹⁸⁴

Auf dem Gebiet der technischen Dokumentation kommen ebenfalls verbreitet *Auszeichnungssprachen* zum Einsatz. DocBook¹⁸⁵ zum Beispiel ist eine Anwendung, die 1991 zur Standardisierung von UNIX-Dokumentationen entwickelt wurde. Sie findet mittlerweile auch auf anderen Gebieten Verwendung.

Da *Auszeichnungssprachen* eine strenge Strukturierung von Texten bedingen, lassen sich entsprechend kodierte Texte gut in Datenbanken, die die Grundlage für elektronische Redaktionssysteme bzw. Content-Management-Systeme bilden, übernehmen. Solche Systeme kommen beispielsweise in großem Umfang in Zeitschriftenverlagen zum Einsatz.

¹⁸² Vgl. 2.2.6 X-Media-Publishing, S. 64.

¹⁸³ Vgl. <http://www.microsoft.com/germany/msdn/office/openxml/default.mspix> (18.05.2008).

¹⁸⁴ Vgl. <http://docs.oasis-open.org/office/v1.1/OS/OpenDocument-v1.1-html/OpenDocument-v1.1.html> (18.05.2008).

¹⁸⁵ <http://www.docbook.org> (18.05.2008).

XML gewinnt darüber hinaus zunehmende Bedeutung im Zusammenhang mit sogenannten objektorientierten Modellen in der Informatik. Objektorientierte Datenbanksysteme (OODBS) und objektorientierte Programmiersprachen (z. B. C++, Java oder NET-Technologien), orientieren sich bei der Modellbildung an der »natürlichen Sichtweise« von Menschen.¹⁸⁶ Hier bilden sich neue Anknüpfungspunkte zwischen Geisteswissenschaft und Informatik.

3.1.6 Auszeichnungssprachen in den Geisteswissenschaften

Wenn man die Aufgaben und Zielsetzungen einer Edition mit den Entwicklungszielen und der praktischen Verwendung von *Auszeichnungssprachen* vergleicht, lassen sich Gemeinsamkeiten feststellen. Eine sehr allgemeine Definition des Gegenstandes der Editionswissenschaft beschreibt ihre Aufgabe als die »Bewahrung von Zeichen und Darstellung derselben in einer Form, die die fortwährende Neuerschließung ihres Sinnes erlaubt«. ¹⁸⁷ Dies kommt der Intention, die zur Entwicklung von *Auszeichnungssprachen* geführt hat, recht nahe. *Auszeichnungssprachen* wie *XML* bieten die Möglichkeit der semantischen Kodierung von Texten. Eine gut konzipierte Textauszeichnung ermöglicht verschiedene Zugänge zum Text: Je nachdem, welche Bedeutung bzw. Funktion man der Auszeichnung zuweist, sind unterschiedliche und ggf. neue Aspekte des Textes darstellbar. Eine »fortwährende Neuerschließung« wird somit unterstützt.

Im Vergleich mit historisch-kritischen Editionen in Buchform liegt der Vorteil einer »Bewahrung« in einem mittels einer *Auszeichnungssprache* realisierten Datenformat darin, dass man sich noch keine Gedanken über die endgültige Form der Darstellung machen muss. Wenn die erforderlichen Informationen in einer geeigneten Daten- bzw. Auszeichnungsstruktur kodiert und gespeichert wurden, lässt sich die Darstellungsform der Fragestellung entsprechend wählen.

Im Gegensatz zu anderen Datenformaten (und manchen historisch-kritischen Ausgaben) sind die meisten *Auszeichnungssprachen* so ausgelegt, dass die Daten auch für einen menschlichen Leser »lesbar« bleiben, d. h. selbst wenn man sich keiner speziellen Software zur Darstellung oder Auswertung des gespeicherten Textes bedient, sind alle Informationen in einem Dokument zu erkennen. Wie komfortabel das möglich ist, hängt von der zugrunde gelegten Datenstruktur ab.

Somit rückt die »Form der Bewahrung« in den Focus der Betrachtung. Nicht die endgültige Form der Darstellung, die spätestens durch den Druck fixiert wird, steht im Mittelpunkt des Interesses, sondern eine möglichst flexible Auszeichnungsstruktur, die dynamisch den jeweiligen Erfordernissen angepasst werden kann.

Wenn man den Stellenwert von *Auszeichnungssprachen* im Umfeld von Technik und Informatik betrachtet, stellt man fest, dass sich die geisteswissenschaftlichen Disziplinen im Vergleich auffallend zurückhalten. Dies ist bedauerlich, da es hier durchaus gegenseitige Interessen gibt. Zum einen bilden *Auszeichnungssprachen* in vielen Fällen bereits das technische Rückgrat geisteswissenschaftlicher Informationssysteme wie Internetressourcen, CD-ROM Editionen oder eBooks und werden auch bei der Erstellung von Publikationen im Rahmen von Datenbanksystemen oder in der Druckvorstufe eingesetzt; zum anderen besteht von Seiten der Informatik Interesse an geisteswissenschaftlichen Theorien und Modellen für Informationsstrukturen und -verarbeitung von Texten.

186 Im Gegensatz zu hierarchischen oder relationalen Datenmodellen bildet ein objektorientiertes Datenmodell einen komplexen Gegenstand als ein Objekt ab. Eigenschaften und Funktionen (Attribute und Methoden) eines Gegenstandes werden nicht, wie z. B. in einem relationalen Modell, in getrennten Bereichen gespeichert, sondern als Einheit betrachtet. Zur Veranschaulichung: In einem relationalen Modell würde eine blaue Kugel dadurch beschrieben, dass eine Relation zu »blau« in einer Liste von Eigenschaften und eine zu »rollen« in einer Liste von Funktionen besteht. Ein objektorientiertes Modell beschreibt eine Kugel etwa in der Form »Kugel, ist blau, kann rollen«.

187 Dieter, Jörg: Historisch-Kritische Edition im virtuellen Raum. Möglichkeiten und Grenzen elektronischer Editionen vor dem Hintergrund einiger Kernprobleme der Editionswissenschaft. 2002. <http://www.jolifanto.de/wissenschaft/webrhetorik/Textedition.pdf> (18.05.2008).

Gründe für diese Zurückhaltung seitens der Geisteswissenschaften mögen in den fachlichen Hürden bei der Verständigung zwischen den beiden Disziplinen »Geist« und »Technik« liegen, welche aber durch einen konstruktiven Austausch zu überwinden sein sollten. Ein weiterer Grund liegt sicher darin, dass sich elektronische Medien – besonders in Bezug auf ihre Verwendung in den Geisteswissenschaften – noch konstituieren und die Entwicklung bislang keinen endgültigen oder absehbaren Stand erreicht hat. Die aufgezeigte Diskrepanz zwischen buchnahen und elektronischen Medien nahen elektronischen Editionen darf nicht als das Ende einer Entwicklung, sondern als Zwischenstand bei der Entwicklung von Konzepten für elektronische Editionen verstanden werden.

Viele Anregungen auf diesem Gebiet kommen aus dem angloamerikanischen Sprachraum, wo das Thema unter dem Schlagwort »Humanities Computing« unter verschiedenen Aspekten diskutiert wird. Wichtige Entwicklungen und Richtlinien sind hier etwa die *Text Encoding Initiative* oder die *Dublin Core Metadata Initiative*. Aufgrund anderer wissenschaftlicher Traditionen, besonders im Bereich der Edition, sind die hier vorgestellten Überlegungen und Ergebnisse nicht uneingeschränkt auf die germanistische Editionswissenschaft zu übertragen. Im deutschen Sprachraum findet die Beschäftigung mit diesem Thema von verschiedenen Seiten statt.

Beachtenswerte Untersuchungen zum Thema elektronischer Texte kommen aus dem Gebiet der Linguistik. Neben einer dem linguistischen Textverständnis entsprechenden Auszeichnung werden vor allen Dingen *Hypertextmodelle* diskutiert.¹⁸⁸ Dabei werden alle Aspekte unter dem Begriff der Texttechnologie zusammengefasst. Bei den verschiedenen diskutierten Kategorien käme die der E-Texte dem Gegenstand dieser Arbeit am nächsten.¹⁸⁹

Da das Interesse, Texte und Quellen in elektronischer Form verfügbar zu machen, über germanistische Belange hinausgeht, sind die Entwicklungen in anderen, ähnlich gelagerten Disziplinen zu beachten. Besonders die Geschichtswissenschaften tun sich hier im Bereich elektronischer Quellenedition hervor. Zurzeit gibt es in Deutschland verschiedene Arbeitsstellen, an denen an der Entwicklung und Umsetzung von *XML*-basiereten Editionsmodellen gearbeitet wird. Die meisten sind aus DFG-Förderungen im Rahmen des »Förderbereichs zum Aufbau einer ›Verteilten Digitalen Forschungsbibliothek« bzw. aus dem Förderprogramm »Retrospektive Digitalisierung von Bibliotheksbeständen« hervorgegangen.¹⁹⁰

3.2 Extensible Markup Language

Um die Konzepte und Beispiele für elektronische Editionen, die ich im Folgenden vorstellen werde, verstehen zu können, sind einige Grundkenntnisse in *XML* erforderlich. Ich beginne diesen Abschnitt deshalb mit einer kurzen Einführung in *XML*. Als Beispiel für eine Anwendung, die für die Auszeichnung von Texten entwickelt wurde, stelle ich daran anschließend die *Text Encoding Initiative* vor.

188 Vgl. Lobin, Henning (Hg.): Text im digitalen Medium. Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering. Wiesbaden: Westdeutscher Verlag, 1999; sowie: Lobin, Henning und Lemnitzer, Lothar (Hg.): Texttechnologie: Perspektiven und Anwendungen. Tübingen: Stauffenburg, 2003.

189 Vgl. Storrer, Angelika: Text und Hypertext. In: Texttechnologie: Perspektiven und Anwendungen. Hg. von Henning Lobin und Lothar Lemnitzer. Tübingen: Stauffenburg, 2003. S. 13-50. Hier besonders S. 28 ff.

190 Seit dem Abschluss des Förderprogramms sind die diesbezüglichen Informationen der DFG leider nicht mehr verfügbar. Eine Liste der geförderten Projekte findet sich unter <http://www.dl-forum.de> (18.05.2008). Vgl. auch Altrichter, Helmut: Retrodigitalisierung in Deutschland – Versuch einer Zwischenbilanz. <http://www.bsb-muenchen.de/mdz/forum/altrichter/> (18.05.2008).

3.2.1 Einführung in XML

Ursprünglich wurde *XML* entwickelt, um eine einfachere, weniger restriktive *Auszeichnungssprache* für Anwendungen im *World Wide Web* zur Verfügung zu haben als *SGML*. Die strengen *SGML*-Vorgaben erlaubten nicht die hier gewünschte Flexibilität. Gleichzeitig wurden einige unübersichtliche, die einfache Auswertung erschwerende Verkürzungsmechanismen, die *SGML* vorsieht, fallen gelassen. Bei der Konzeption von *XML* wurden auch neuere Entwicklungen in der Zeichenkodierung berücksichtigt. *XML* verwendet standardmäßig den *Unicode*-Zeichensatz und umgeht so einige Probleme, die sich bei einer *SGML*-Auszeichnung von Texten auf der Ebene des hier zugrunde gelegten *ASCII*-Kodes ergeben.

Wenn im Zusammenhang mit *XML* und *SGML* die Rede von »Sprachen« ist, so ist dies streng genommen nicht korrekt. *SGML* und *XML* sind keine Sprachen im eigentlichen Sinne, sondern Metasprachen. Sie sind als eine Art Grammatik zu verstehen, auf deren Grundlage konkrete *Auszeichnungssprachen* definiert werden können. Zur Beschreibung von Metasprachen wird oft das Modell der Formalen Sprachen Chomskys herangezogen.¹⁹¹ *SGML* bzw. *XML* können nach diesem Modell als einfache, kontextfreie (CH-2)-Grammatiken betrachtet werden. Das heißt, die Metasprachen *SGML* bzw. *XML* beschreiben mögliche *Auszeichnungssprachen*.

Der Begriff »Auszeichnung« wird in Bezug auf Texte meist für Anmerkungen oder Markierungen verwendet, die in einem Text, etwa einem Manuskript, angebracht werden, um einem Setzer oder Drucker mitzuteilen, wie eine bestimmte Textstelle gedruckt oder gestaltet werden soll. Dazu müssen Regeln aufgestellt werden, die festlegen, welche Auszeichnungen es gibt, wie sie verwendet werden und was sie bedeuten.

SGML und *XML* beschreiben, welche Auszeichnungen es gibt und wie sie verwendet werden. Auszeichnungen sind rein deskriptiv und dazu ausgelegt, Strukturen in einem Dokument zu beschreiben. Was die Auszeichnungen genau bedeuten, kann zwar etwa durch die Namensgebung der Auszeichnungsmarkierungen angegeben werden, hat aber keine Funktion. *SGML/XML*-konforme Auszeichnungen sind nicht verfahrensorientiert, d. h. sie beschreiben nicht, was letztendlich mit dem markierten Text zu geschehen hat.

Zur Beschreibung der Textstruktur bietet *XML* zwei Stufen an. Die erste berücksichtigt nur sehr allgemeine Regeln zur Textauszeichnung. Sie wurde eingeführt, um mit wenig Aufwand Texte für einfache Anwendungen kodieren zu können. Interessant für komplexere Anwendungen ist hauptsächlich die zweite Stufe. Sie verlangt, dass einem entsprechend ausgezeichneten Text ein konkretes *Inhaltsmodell* zugrunde gelegt wird. Das *Inhaltsmodell* dient dazu, die Struktur jedes nach diesem Modell kodierten Textes vorzugeben. Darüber hinaus wird im *Inhaltsmodell* das »Vokabular« der konkreten Anwendung bzw. *Auszeichnungssprache* definiert.

3.2.2 XML Grundbegriffe

An dieser Stelle möchte ich einen kleinen Überblick über die wichtigsten Grundbegriffe und Regeln von *XML* geben. Ich werde nur die Grundlagen referieren, auf die ich im Folgenden zurückgreife und die zum Verständnis der weiteren Überlegungen notwendig sind.¹⁹²

Einzelheiten und die genauen Spezifikationen von *XML* finden sich im Internet auf den Seiten des World Wide Web Consortiums.¹⁹³

191 Der theoretische Hintergrund aus der Informatik sind »formale Sprachen«. Dieser Begriff geht auf Noam Chomsky zurück, der bereits in den 1950er Jahren versucht hat, natürliche Sprache mit formalen Modellen zu beschreiben. In der Informatik führten diese Arbeiten zu mehreren Modellen (»Chomsky-Hierarchie« der Grammatiken).

192 Mittlerweile ist die Literatur zum Thema XML unüberschaubar. Die Entwicklung von XML wird vom World Wide Web Consortium (W3C) koordiniert. Hier finden sich auch die aktuellsten Informationen zum Thema: <http://www.w3.org> (18.05.2008). Eine sehr gute Einführung in XML in englischer Sprache ist das Kapitel 2 der TEI-Guidelines, das auch separat unter <http://www.tei-c.org/Guidelines2/gentleintro.xml> einsehbar ist.

193 <http://www.w3.org/XML/> (18.05.2008).

Tags

Wie ich oben erwähnt habe, sind die Regeln, die *XML* vorgibt, dazu da, eine konkrete *Auszeichnungssprache* erst zu definieren. Das erste, was zur Auszeichnung von Texten überhaupt festgelegt werden muss, ist die Art, wie die Auszeichnungen vorgenommen werden. Berücksichtigt man, wie Computer arbeiten, muss zuerst definiert werden, wie Auszeichnungen von ausgezeichnetem Text unterschieden werden.¹⁹⁴ *XML* gibt hier lediglich vor, dass der Beginn einer Auszeichnungsmarkierung durch eine öffnende spitze Klammer bzw. das Zeichen »kleiner als« » < « gekennzeichnet wird. Dann folgt ein frei wählbarer Name für die Auszeichnung, der mit einem Buchstaben beginnen muss, im Weiteren jedoch aus beinahe allen beliebigen Zeichen und Zeichenfolgen bestehen darf. Nicht erlaubt sind lediglich einige festgelegte Schlüsselwörter und -zeichen sowie sogenannte »Whitespaces«, das sind Leerzeichen, Zeilenumbrüche, Absatzmarken oder Tabulatoren, da diese zur Trennung verschiedener Angaben in *XML* dienen.¹⁹⁵ Das Ende einer Auszeichnungsmarkierung wird durch eine schließende spitze Klammer bzw. das Zeichen »größer als« » > « gekennzeichnet. Eine komplette Auszeichnungsmarkierung wird als *Tag* bezeichnet:

```
<Tag>
```

Neben dem Beginn eines markierten Bereiches muss sein Ende ausgezeichnet werden. Dazu wird der gleiche *Tag* verwendet, der den Anfang des markierten Bereiches auszeichnet. Auf die öffnende spitze Klammer folgt bei einer Endmarkierung zunächst ein Schrägstrich » / «. Da die beiden *Tags* den markierten Text umklammern, spricht man, analog zur normalen Verwendung von Klammern, nicht von Anfangs- und End-*Tags* sondern von öffnenden und schließenden *Tags*:

```
öffnender Tag:      <Tag>
schließender Tag:  </Tag>.
```

Elemente

Durch einen öffnenden und einen schließenden *Tag* markierte Bereiche bilden die Grundeinheiten oder auch Struktureinheiten von *XML*. Eine solche Einheit wird **Element** (element) genannt. Den für die Auszeichnungsmarkierung – also den *Tag* – vergebenen Namen bezeichnet man als **Elementnamen** (generic identifier, GI); den durch die *Tags* markierten Text als **Inhalt** (content):

```
Element:      <Name>Inhalt</Name>
```

An dieser Stelle wird deutlich, was damit gemeint ist, wenn *XML* als Metasprache bezeichnet wird. Festgelegt wird lediglich die Form der Auszeichnung. Wie die Auszeichnungen heißen und was ausgezeichnet wird, bleibt zunächst offen. Da Elementnamen frei wählbar sind, bietet es sich an, aussagekräftige Namen zu verwenden.

Der Inhalt eines Elements kann sowohl aus Text als auch aus weiteren Elementen bestehen. Was ein Element genau beinhalten darf, kann durch ein *Inhaltsmodell* bestimmt werden.¹⁹⁶

Elemente, die keinen Inhalt haben, bezeichnet man als **leere Elemente**. Sie bestehen nur aus einem *Tag*, bei dem der Schrägstrich vor der schließenden spitzen Klammer steht.

```
<Leeres_Element/>.
```

Instanzen

Elemente werden benutzt, um Inhalte gleicher Art auszuzeichnen. Verschiedene Elemente der gleichen Art bezeichnet man auch als Elemente der gleichen Gattung (daher auch der englische Begriff »generic identifier« – Gattungsbezeichner – für den Elementnamen). Elemente einer Gattung können, mit Ausnahme des

194 Vgl. 2.2.3 Kodierung elektronischer Texte, S. 60.

195 In Bezug darauf, welche Zeichen für Elementnamen verwendet werden dürfen, unterscheiden sich die *XML*-Versionen. *XML* 1.0 und 1.1.

196 Siehe 3.2.4 *XML* Inhaltsmodelle, S. 87.

Wurzelements (s. u.), beliebig oft in einem Dokument vorkommen. Eine Liste von Autoren könnte etwa folgendermaßen aussehen:

```
<Autorenliste>
  <Autor>Thomas Mann</Autor>
  <Autor>Georg Büchner</Autor>
  <Autor>Heinrich Heine</Autor>
  <Autor>Friedrich Schiller</Autor>
</Autorenliste>
```

Beispiel 3. Autorenliste

Das Element »Autor« kommt hier vier Mal in dem Element »Autorenliste« vor. Um die Gattung der Elemente mit dem Namen »Autor« von dem konkreten Vorkommen des Elements mit einem bestimmten Inhalt zu unterscheiden, spricht man im letzteren Falle von einer Instanz (instance) des Elements.

```
<Autor>Thomas Mann</Autor>
```

ist also eine Instanz des Elements »Autor«. Eine Instanz kann wiederum verschiedene Instanzen beinhalten: Die oben dargestellte Instanz des Elements »Autorenliste« enthält vier Instanzen des Elements »Autor«. Was im Folgenden über Elemente gesagt wird, gilt genauso für die jeweils konkreten Instanzen des Elements.

Struktur

Neben der Form der Auszeichnung gibt *XML* auch konkrete Strukturen vor, der die Auszeichnungen zu folgen haben. Die wichtigste Regel dabei ist, dass Elemente sich nicht überschneiden dürfen. Elemente können verschachtelt werden (innerhalb eines Elements können beliebig viele weitere Elemente vorkommen), ein untergeordnetes Element muss aber immer vollständig in dem jeweils übergeordneten Element enthalten sein.

Die oben angeführte Struktur ist demnach richtig:

```
<Äußerung>
  <Sprecher>Danton</Sprecher>
  <Text>Spazieren, mein Junge, spazieren.
  <Regieanweisung>Er geht.</Regieanweisung>
</Text>
</Äußerung>
```

Beispiel 4. Zeile aus DANTONS TOD semantisch ausgezeichnet

Falsch bzw. nicht *XML*-konform wäre dagegen folgende Struktur:

```
<Äußerung>
  <Sprecher>Danton</Sprecher>
  <Text>Spazieren, mein Junge, spazieren.
  <Regieanweisung>Er geht.
  </Text>
</Regieanweisung>
</Äußerung>
```

Beispiel 5. Schachtelungsfehler

Das Element »Regieanweisung« wird innerhalb des Elements »Text« geöffnet, aber außerhalb geschlossen. Auch wenn dieses Beispiel etwas willkürlich erscheint, gibt es durchaus Fälle, in denen eine Struktur, in der sich Elemente überschneiden, angebracht erscheinen kann, etwa wenn sich verschiedene Strukturen, die berücksichtigt werden sollen, überlagern. Zum Beispiel kann es wünschenswert sein, neben der Kapitelstruktur eines Romans auch die Seitenstruktur einer Buchvorlage zu berücksichtigen:

```

<Seite Nummer="1">197
  <Kapitel>
    <Text>...
  </Seite>
<Seite Nummer="2">
  </Text>
</Kapitel>

```

Beispiel 6. Schachtelungsfehler

Eine weitere Regel, die von *XML* vorgegeben wird, ist, dass es immer ein Element geben muss, das alle anderen Elemente einschließt, das sogenannte **Wurzelement** (root element). Im Beispiel 4 ist das Element »Äußerung« das Wurzelement. Bei mehreren Äußerungen müsste das ganze von einem Element, das »Szene« heißen könnte, umschlossen werden. Enthält das Dokument mehrere Szenen, müsste wieder ein Wurzelement, etwa »Akt« ergänzt werden, das alle Szenen beinhaltet. Mehrere Akte müssten wieder unter ein Element wie »Drama« fallen, das alle Teile des Dramas umschließt, also ggf. auch ein Element »Personenverzeichnis«, ein Element »Prolog« und so weiter.

Entspricht ein *XML*-Dokument diesen einfachen Regeln, wird es als **wohlgeformt** (well formed) bezeichnet. Damit ein *XML*-Dokument auch den strengeren *XML*-Spezifikationen entspricht, benötigt es ein *Inhaltsmodell*.

Attribute

Attribute (attributes) bieten die Möglichkeit, die Auszeichnungen um Informationen zu ergänzen, die nicht unmittelbar zum Text gehören. Attribute können in den öffnenden *Tag* eines Elements oder in den *Tag* leerer Elemente eingefügt werden. Ein Attribut besteht aus dem **Attributnamen** (attribute name) und dem **Attributwert** (attribute value). Der Attributwert steht in einfachen oder doppelten Anführungszeichen und wird dem Attributnamen durch ein Gleichheitszeichen zugeordnet:

```
<Element Attribut="Wert"> ... </Element>
```

Ein Element kann um ein oder mehrere Attribute ergänzt werden.

Entitäten

Entitäten (entities) sind Stellvertreter für Informationen, die Teil eines Dokuments sind, aber an einer anderen Stelle festgelegt wurden. Dies kann ein einzelnes Zeichen oder auch ein ganzes Dokument sein. Entitäten beginnen mit einem »Kaufmanns und Zeichen« & « und enden mit einem Semikolon. Dazwischen steht ein beliebiger Name:

```
&Entität;
```

Entitäten können beispielsweise benutzt werden, um einzelne Zeichen zu repräsentieren. Da das »kleiner als Zeichen und das »Kaufmanns und« in *XML*-Texten eine besondere Bedeutung haben (» < « leitet einen *Tag* ein, » & « eine Entität), dürfen sie im sonstigen Text nicht vorkommen. Sie müssen deshalb in allen *XML*-Dokumenten durch Entitäten ersetzt werden:

```
&l t; ersetzt „<“ (lower than)
```

```
&a m p; ersetzt „&“ (ampersand)
```

Weitere Einsatzmöglichkeiten von Entitäten bestehen etwa darin, mehrmals vorkommende, identische Passagen durch eine Entität zu ersetzen oder Entitäten als Variablen zu benutzen, die mit verschiedenen Inhalten aufgefüllt werden können.

¹⁹⁷ Um den Sinn einer Seitenmarkierung zu veranschaulichen, habe ich das Element »Seite« um ein Attribut ergänzt – s. u.

In den meisten Fällen muss eine Entität definiert werden. Sofern eine Entität nicht in *XML* selber definiert ist, wie beispielsweise »<« und »>«, geschieht dies innerhalb der *Dokument Type Definition*¹⁹⁸ oder eines anderen *Inhaltsmodells*. Dabei kann eine entsprechende Definition die Zeichen enthalten, die durch die Entität repräsentiert werden sollen oder auch einen Verweis auf den Ort, an dem diese Zeichen – ggf. ein ganzes Dokument – hinterlegt sind.

CDATA-Abschnitte

Wie ich oben dargestellt habe, dürfen in *XML*-Dokumenten bestimmte Zeichen, etwa die Zeichen, die bestimmte Funktionen haben, wie die spitzen Klammern oder das »Kaufmanns und«, nicht vorkommen. Daneben müssen alle Teile des Dokuments *XML*-konform sein, d. h. den Schachtelungsregeln entsprechen. Um Inhalte in *XML*-Dokumente einbinden zu können, die nicht *XML*-konform sind, beispielsweise Daten, die nach einem anderen Standard kodiert sind, bietet *XML* die Möglichkeit, sogenannte CDATA, character data bzw. Zeichendaten einzufügen. Der Beginn eines solchen Abschnittes wird durch die Zeichenfolge

```
<![CDATA[
```

eingeleitet und durch

```
]]>
```

beendet. Alle Zeichen, die dazwischen stehen, werden von Programmen, die mit *XML* arbeiten, ignoriert, können aber ggf. von anderen Programmen ausgewertet werden.

Prolog

Am Beginn eines *XML*-Dokuments steht ein sogenannter Prolog. Im einfachsten Fall besteht dieser lediglich aus der *XML*-Deklaration, die zum einen besagt, dass es sich im Folgenden um ein *XML*-Dokument handelt, welche *XML*-Version dem Dokument zugrunde gelegt wurde und welcher Zeichensatz dem Dokument zugrunde liegt:

```
<?xml version="1.0" encoding="UTF 8" ?>
```

Der Prolog kann auch umfangreicher ausfallen, etwa, wenn er ein *Inhaltsmodell* enthält oder auf eines verweist, wenn *Namespaces* in ihm definiert werden etc.

Inhaltsmodell

Ein Inhaltsmodell (content model) beschreibt, wie das Dokument aufgebaut ist. Im Wesentlichen wird hier definiert, welche Elemente wie oft und ggf. in welcher Reihenfolge vorkommen dürfen, wie sie verschachtelt sein dürfen und was sie enthalten können. Das *Inhaltsmodell* kann in einem separaten Abschnitt im Dokumentkopf dem eigentlichen Dokument vorangestellt oder in eine externe Datei ausgelagert werden. Üblicherweise verwendet man zur Beschreibung des *Inhaltsmodells* eine sogenannte **Dokumenttyp Definition** (*document type definition*), kurz *DTD*. Für den Aufbau und die Struktur von *DTDs* gibt es eigene Regeln. Neben den vorkommenden Elementen werden in einer *DTD* auch Attribute und Entitäten definiert. Soll einem *XML*-Dokument ein *Inhaltsmodell* zugrunde gelegt werden, müssen dem eigentlichen Dokument bzw. dem Dokumentinhalt noch entsprechende Informationen vorangestellt werden. Dies kann die vollständige *DTD* sein – man spricht dann von einer internen *DTD* – oder der Verweis auf den Ort, an dem eine externe *DTD* gespeichert ist.

Ein *XML*-Dokument, für das ein entsprechendes *Inhaltsmodell* bzw. eine *DTD* existiert und das damit abgeglichen, **validiert** (validated), werden kann, bezeichnet man als **gültig** (valid).

Programme, die auf ein gültiges *XML*-Dokument zugreifen, bekommen durch das Einlesen der *DTD* Informationen über Struktur und inhaltliche Aspekte des Dokuments. Eine weitergehende Spezifikation von Elementinhalten kann über andere Modelle zur Inhaltsbeschreibung, beispielsweise ein **XML-Schema** erfolgen.

198 Siehe 3.2.4 XML-Inhaltsmodelle, S. 87.

Ein Schema erlaubt es, den Inhalt von Elementen genauer zu bestimmen. Kann in einer *DTD* lediglich angegeben werden, wie oft und in welcher Reihenfolge ein Element andere Elemente oder beliebigen Text beinhalten darf, kann mit Hilfe eines *XML-Schemas* festgelegt werden, wie der Text, den ein Element enthält, aussehen muss. Etwa, wie lang er sein muss, ob er aus Buchstaben oder Zahlen bestehen muss, ob die Zahlen eine bestimmte Struktur, etwa die eines Datums aufweisen müssen. Ein *XML*-Dokument, das gegen ein Schema abgeglichen werden kann, bezeichnet man als **schemagültig** (schema-valid).

Whitespaces

Das Wort »Whitespace« – weißer Platz bzw. Raum – bezeichnet alle Zeichen, die dazu dienen, Platz zwischen Buchstaben zu schaffen. Dazu zählen einfache Leerzeichen (engl. »space«) genauso wie Tabulatoren, Zeilen oder Absatzumbrüche. Sie werden im Rahmen von *XML* meist gleich behandelt und dienen als Hilfsmittel zur Darstellung der Struktur. Die meisten Programme, die *XML*-Daten verarbeiten, stellen Whitespaces als ein Leerzeichen dar. Auch mehrere aufeinander folgende Leerzeichen werden als ein Leerzeichen ausgegeben.

Folgende *XML*-Strukturen sind deshalb identisch:

```
<Text> <Absatz> <Überschrift> Titel </Überschrift> <Zeile> Inhalt </Zeile>
<Zeile> Inhalt </Zeile> </Absatz> </Text>

<Text>
  <Absatz>
    <Überschrift>
      Titel
    </Überschrift>
    <Zeile>
      Inhalt
    </Zeile>
    <Zeile>
      Inhalt
    </Zeile>
  </Absatz>
</Text>
```

Beispiel 7. Strukturdarstellung mit Whitespaces

Der Vorteil der unteren Darstellungsweise ist, dass alle Elemente, die sich auf einer Strukturebene befinden, untereinander stehen. Strukturfehler sind so leicht erkennbar. Sollen innerhalb eines Textes Abstände, Zeilenumbrüche oder zweidimensionale Strukturen (z. B. Tabellen) dargestellt werden, muss dies speziell kodiert werden.

3.2.3 Darstellungsweisen

Zum einfacheren Verständnis von *XML*-Strukturen, auf die weitere *XML*-Technologien aufbauen, möchte ich noch zwei gängige Darstellungsweisen von *XML*-Dokumenten vorstellen.

Die Elementinstanzen können als Kreise dargestellt werden. Diese Darstellung ähnelt Darstellungen der mathematischen Mengenlehre. Da es nach den Regeln von *XML* nicht erlaubt ist, dass Elemente sich überschneiden, darf es keine Schnittmengen geben.

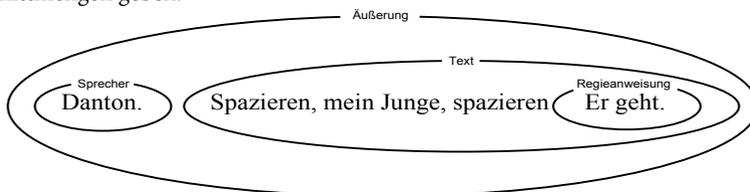


Abbildung 21. Darstellung der XML-Struktur in Anlehnung an die Mengenlehre

Diese Darstellung wird zur Veranschaulichung der Struktur mit einem Schwerpunkt auf den Inhalten verwendet, etwa bei der Analyse zu strukturierender Dokumente.

Die zweite, weiter verbreitete Darstellungsweise ist die Baumstruktur. Hier liegt der Schwerpunkt der Abbildung auf der Struktur. Die Baumstruktur bildet die Grundlage für die meisten Navigationstechnologien, wie *XPath*¹⁹⁹ etc.:

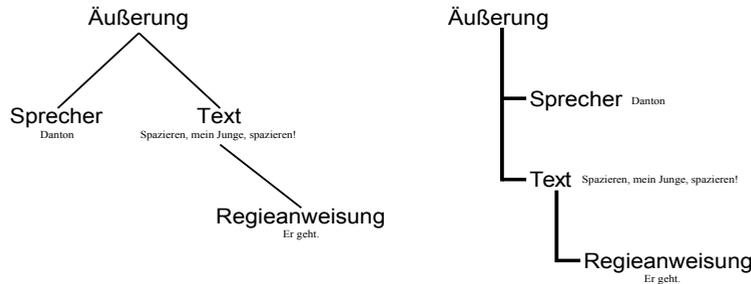


Abbildung 22. Darstellung der XML-Struktur als Strukturbaum

3.2.4 XML-Inhaltsmodelle

Wie oben dargelegt, besteht die Möglichkeit, für ein *XML*-Dokument oder eine *XML*-Anwendung ein *Inhaltsmodell* zu erstellen. Zur Angabe eines *Inhaltsmodells* sind verschiedene Technologien entwickelt worden. Sie ermöglichen es, *Inhaltsmodelle* unterschiedlich konkret bzw. komplex anzugeben. Die drei wichtigsten stelle ich hier kurz vor. Die *DTD* stellt die im Vergleich zu den anderen vorgestellten Technologien älteste und einfachste Möglichkeit dar, ein *Inhaltsmodell* für eine *XML*-Anwendung zu erstellen. *XML*-Schema und *RELAX-NG* sind neuere Entwicklungen die unterschiedliche Möglichkeiten bei der Definition eines *Inhaltsmodells* bieten.

DTD

Ursprünglich war für *XML*-Dokumente vorgesehen, ein *Inhaltsmodell* mit Hilfe einer sogenannten *DTD* zu erstellen. *DTDs* werden innerhalb der *SGML*/*XML*-Spezifikationen beschrieben. In einer *DTD* wird festgelegt, welche Elemente in einem *XML*-Dokument vorkommen dürfen oder müssen, wie sie verschachtelt sein können oder sollen, welches Element welche *Attribute* hat und welche Entitäten es gibt und wofür sie stehen. Eine *DTD* dient zum einen dazu, ein *XML*-Dokument zu validieren und damit zu kontrollieren, ob es den Vorgaben des *Inhaltsmodells* entspricht. Ist dies der Fall, spricht man von einem »gültigen« Dokument. Zum anderen liefert eine *DTD* einem Programm, das auf ein *XML*-Dokument zurückgreift, Informationen über die Struktur des Dokuments.

Eine *DTD* wird in einer eigenen Syntax verfasst. Ich verzichte im Rahmen dieser Arbeit darauf, das Thema zu vertiefen. Für die meisten *XML*-Anwendungen, die ich im Folgenden vorstellen werde, stehen *DTDs* zur Verfügung oder es werden andere Technologien zur Beschreibung des *Inhaltsmodells* benutzt.

Auf der einen Seite ist der Umgang mit *DTDs* schwierig und komplex, auf der anderen Seite ist es in vielen Fällen wünschenswert, ein *Inhaltsmodell* noch konkreter zu formulieren, als es mit einer *DTD* möglich ist. Aus diesem Grund wurden weitere Schemasprachen als Alternativen zur *DTD* entwickelt.

199 Vgl. Technologien zur Adressierung, S. 93.

XML-Schema

Ein *XML-Schema* (XSD) ermöglicht eine detailliertere Beschreibung eines Inhaltsmodells als eine *DTD*. Während sich eine *DTD* lediglich dazu eignet, grob die Dokumentstruktur zu beschreiben, erlaubt sie es beispielsweise nicht, komplexere Reihenfolgen von Elementen bzw. deren Vorkommen vorzugeben oder die Zeichendaten, die ein Element oder ein Attribut enthalten darf, näher zu bestimmen. Hier sind die Möglichkeiten von *XML-Schema* deutlich komplexer. Ein *XML-Schema* ist selber eine *XML*-Anwendung, d. h. es handelt sich um ein *XML*-Dokument. Um Konflikte mit der *XML*-Anwendung, die mittels des *XML-Schemas* beschrieben werden soll, zu vermeiden, wird auf Namespaces²⁰⁰ zurückgegriffen. Damit unterstützt *XML-Schema* im Gegensatz zur *DTD* die Verwendung von Namensräumen. Ein konkretes *XML-Schema* wird als XML-Schema-Definition (XSD) bezeichnet.

Weitere Vorteile sind, dass ein *XML-Schema* es erlaubt, genau die Datentypen zu definieren, die ein Element enthalten darf und geht damit weit über die Möglichkeiten einer *DTD* hinaus. Während man beispielsweise den Inhalt eines Elements, das ein Datum enthalten soll, mit Hilfe einer *DTD* als CDATA, also »Zeichendaten« deklarieren kann, erlaubt ein *XML-Schema* die genaue Angabe, dass das entsprechende Element ein Datum beinhalten soll und in welcher Form es angegeben werden muss.

XML-Schema ist die zurzeit vom World Wide Web Consortium empfohlene Technologie um *Inhaltsmodelle* für *XML*-Dokumente zu definieren.²⁰¹

RELAX-NG

Eine andere Entwicklung zur Beschreibung von *XML-Inhaltsmodellen* ist *RELAX-NG*. Die *Regular Language Description for XML New Generation* ist eine weitere Schemasprache für *XML*. Im Gegensatz zu *XML-Schema* wird sie nicht vom W3C gepflegt, sondern von der »Organization for the Advancement of Structured Information Standards« (OASIS),²⁰² die auch den Doc-Book Standard und den Open-Document Standard betreut.²⁰³

RELAX-NG stellt eine Grammatik zur Beschreibung der Struktur von *XML*-Dokumenten dar. Ein *RELAX-NG* Schema kann *XML*-konform als *XML*-Dokument verfasst werden; daneben existiert auch eine alternative, nicht *XML*-konforme Notation.

RELAX-NG beschränkt sich auf die Beschreibung der Struktur eines *XML*-Dokuments und ist damit eher als Alternative zu einer *DTD* zu betrachten. Die konkrete Definition von Inhalten, wie sie ein *XML-Schema* ermöglicht, erlaubt *RELAX-NG* nicht. Dafür sieht es *RELAX-NG* vor, auf *XML-Schema* zurückzugreifen, wenn dies erforderlich ist und beide Technologien zu kombinieren.

Der Vorteil von *RELAX-NG* besteht darin, dass sie einerseits weniger restriktiv ist als *XML-Schema*, andererseits flexibler. *RELAX-NG* wird deshalb gern in Bereichen eingesetzt, in denen eine weniger streng geregelte *XML*-Struktur wichtiger ist als die konkrete Deklaration der Inhalte. Das *Inhaltsmodell* der aktuellen GUIDELINES FOR ELECTRONIC TEXT ENCODING AND INTERCHANGE der TEXT ENCODING INITIATIVE²⁰⁴ wurde mit Hilfe von *RELAX-NG* definiert.

3.2.5 XML-Namensräume

Ein gravierender Nachteil der Beschreibung eines *Inhaltsmodells* mittels einer *DTD* ist, dass einem Dokument nur eine *DTD* zugeordnet werden darf. Ursprünglich sollte damit eine eindeutige Deklaration von Elementen und Entitäten gewährleistet ist. In der Praxis hat es sich jedoch als wünschenswert herausgestellt, dass verschie-

200 Vgl. 3.2.5 XML-Namensräume, S. 88.

201 <http://www.w3.org/XML/Schema> (18.05.2008).

202 <http://www.oasis-open.org/home/index.php> (18.05.2008).

203 3.2.6 XML-Anwendungen, S. 91.

204 Vgl. 3.3 Die Text Encoding Initiative, S. 94.

dene XML-Anwendungen bzw. Technologien in einem Dokument kombiniert werden dürfen. Für einige XML-Technologien, beispielsweise der *Extensible Stylesheet Language*, ist dies sogar Voraussetzung. Als Ausweg wurden XML-Namensräume (*Namespaces*) eingeführt.

Werden zwei XML-Anwendungen in einem Dokument kombiniert, kann es beispielsweise zu Problemen kommen, wenn beide Anwendungen den gleichen Elementnamen für unterschiedliche Elemente verwenden. So schlagen etwa die Richtlinien der *Text Encoding Initiative* vor, dass das Element »title« möglichst den vollständigen Titel eines Dokumentes enthalten soll. Zur näheren Bestimmung des Titels stehen die *Attribute* »level« und »type« zur Verfügung. Die Richtlinien von *The Dublin Core Metadata Initiative*²⁰⁵ geben dagegen vor, dass das Element »title« den Titel des Dokuments, unter dem es »formell bekannt« gemacht wird, enthält, und sieht keine *Attribute* vor. Soll ein Dokument nun nach *TEI* strukturiert werden, bibliografische Angaben innerhalb des Dokuments aber nach *DCMI* erfolgen, muss unterschieden werden, ob ein im Dokument vorkommendes Element »title« nun den Vorgaben von *TEI* oder *DCMI* entsprechen soll.

Namensräume dienen dazu, zu kennzeichnen, woher, aus welcher XML-Anwendung die verwendeten Elementnamen kommen. Als »Referenz« wird jeweils die der *Uniform Resource Locator (URL)*, die »Internet-Adresse« der Anwendung angegeben, der die Elementnamen entnommen sind. Die Angabe der *URL* hat keine tiefergehende Bedeutung. Es muss lediglich gewährleistet werden, dass die Referenz eindeutig ist. Da jede *URL* nur einmal existiert, wird dieser Forderung entsprochen.

Zur Kennzeichnung von *Namensräumen* gibt es zwei Möglichkeiten: Standardnamensräume oder Präfixe. Standardnamensräume gelten für das Element, in dem der jeweilige *Namensraum* definiert wurde, sowie für diesem untergeordnete Elemente (*Child-Elements*). Präfixe müssen vor ihrer Verwendung definiert werden. Dabei ist es egal, ob die Definition in einem übergeordneten Element des Dokuments oder im ersten Element des *Namensraums* (das dann auch das erste Element mit Präfix ist), stattfindet.

Die Beispiele zeigen den Unterschied zwischen einem Standard-*Namensraum* und einem *Namensraum* mit Präfix anhand von Wohnungsanzeigen.

A Friedrichshain: Warschauer Straße: u.a.: Ruhige 1-Zimmer-Altbauwohnung ...

Wohnung	100,80 €	1	24 m ²
Berlin Friedrichshain	NKM zzgl. NK		Wohnfläche

Kreuzberg: Böckstraße im Graefekiez: Erstbezug nach Sanierung. Ruhige ...

Wohnung	270,00 €	1	40 m ²
Berlin Kreuzberg	NKM zzgl. NK		Wohnfläche

Tagesspiegel —

1	san. AB/Schloss Charl.	35	280,-k	zzgl.	+ ggf Einbauküche, gefBd+Fenster, Laminat, Keller,	www.guchsesne-immobilien.de
---	------------------------	----	--------	-------	--	-----------------------------

Friedrichshain: Scharnweberstraße: u.a.: 1-Zimmer-Altbauwohnung an ...

Wohnung	97,52 €	1	24,38 m ²
Berlin Friedrichshain	NKM zzgl. NK		Wohnfläche

Abbildung 23. Beispiel für die Verwendung eines Standardnamensraums

205 Vgl. Die Dublin Core Metadata Initiative, S. 133.

Abbildung 23 zeigt vier Wohnungsanzeigen. Drei davon sind der Berliner Morgenpost entnommen. Neben der dritten ist vermerkt, dass sie aus einer anderen Zeitung, dem Berliner Tagesspiegel, stammt. Die Angaben gehören in einen anderen *Namensraum*, dem des Tagesspiegels. Ein entsprechendes XML-Dokument sähe so aus:

```
<?xml version="1.0" encoding="UTF 8"?>
<Morgenpost>
  <Anzeige>
    <Stadtteil>Friedrichshain</Stadtteil>
    <Zimmer>1</Zimmer>
    <Wohnfläche>24</Wohnfläche>

    <Miete>100,80</Miete>
  </Anzeige>
  <Anzeige>
    <Stadtteil>Kreuzberg</Stadtteil>
    <Zimmer>1</Zimmer>
    <Wohnfläche>40</Wohnfläche>
    <Miete>270</Miete>
  </Anzeige>
  <Anzeige xmlns="http://www.tagesspiegel.de">
    <Stadtteil>Charlottenburg</Stadtteil>
    <Zimmer>1</Zimmer>
    <Wohnfläche>35</Wohnfläche>
    <Miete>280</Miete>
  </Anzeige>
  <Anzeige>
    <Stadtteil>Friedrichshain</Stadtteil>
    <Zimmer>1</Zimmer>
    <Wohnfläche>24,38</Wohnfläche>
    <Miete>97,52</Miete>
  </Anzeige>
</Morgenpost>
```

Beispiel 8. Beispiel für einen Standardnamensraum

Abbildung 24 zeigt, wie Kennzeichnung eines *Namensraums* mit Hilfe eines Präfixes erfolgt (Beispiel 9):

A Friedrichshain: Warschauer Straße: u.a.: Ruhige 1-Zimmer-Altbauwohnung ...						
Wohnung	100,80 €	1	24 m ²			
Berlin Friedrichshain	NKM zzgl. NK		Wohnfläche			
Kreuzberg: Böckhstraße im Graefekiez: Erstbezug nach Sanierung. Ruhige ...						
Wohnung	270,00 €	1	40 m ²			
Berlin Kreuzberg	NKM zzgl. NK		Wohnfläche			
1	ts	san. AB/Schloss Charl.	ts	ts	260,-€	zzgl. * ggf. Einbauküche, geflBd+Fenster, Laminat, Keller, www.duchesne-immobilien.de
Friedrichshain: Scharnweberstraße: u.a.: 1-Zimmer-Altbauwohnung an ...						
Wohnung	97,52 €	1	24,38 m ²			
Berlin Friedrichshain	NKM zzgl. NK		Wohnfläche			

ts = Angaben aus dem Tagesspiegel

Abbildung 24. Beispiel für die Verwendung eines Namensraums mit Präfix

```

<?xml version="1.0" encoding="UTF 8"?>
<Morgenpost xmlns:ts="http://www.tagesspiegel.de">
  <Anzeige>
    <Stadtteil>Friedrichshain</Stadtteil>
    <Zimmer>1</Zimmer>
    <Wohnfläche>24</Wohnfläche>
    <Miete>100,80</Miete>
  </Anzeige>
  <Anzeige>
    <Stadtteil>Kreuzberg</Stadtteil>
    <Zimmer>1</Zimmer>
    <Wohnfläche>40</Wohnfläche>
    <Miete>270</Miete>
  </Anzeige>
  <Anzeige>
    <ts:Stadtteil>Charlottenburg</ts:Stadtteil>
    <ts:Zimmer>1</ts:Zimmer>
    <ts:Wohnfläche>35</ts:Wohnfläche>
    <ts:Miete>280</ts:Miete>
  </Anzeige>
  <Anzeige>
    <Stadtteil>Friedrichshain</Stadtteil>
    <Zimmer>1</Zimmer>
    <Wohnfläche>24,38</Wohnfläche>
    <Miete>97,52</Miete>
  </Anzeige>
</Morgenpost>

```

Beispiel 9. Beispiel Namensraum mit Präfix realisiert

3.2.6 XML-Anwendungen

Unter *XML*-Anwendungen versteht man allgemein konkrete Anwendungen zur Text- bzw. Datenauszeichnung mit *XML*. Wie oben gesagt,²⁰⁶ handelt es sich bei *XML* um eine Metasprache, eine Grammatik, auf deren Grundlage spezifische *Auszeichnungssprachen* entwickelt werden können. Wie die Vorgaben für eine *XML*-Anwendung aussehen, ist dabei zunächst offen. Im einfachsten Fall reicht eine Empfehlung, was überhaupt ausgezeichnet werden soll. Konkreter wäre die Vorgabe, welche Elemente erlaubt sind, also die Festlegung des Vokabulars. Bis hierhin lässt sich ein Dokument, das einer solchen Vorgabe folgt, lediglich auf seine Wohlgeformtheit hin überprüfen. Will man überprüfen, ob ein Dokument den Vorgaben, die die Anwendung macht, entspricht, benötigt man ein geeignetes *Inhaltsmodell*, etwa eine *DTD*. Für die meisten etablierten *XML*-Anwendungen wurden geeignete *DTDs* entwickelt. Noch konkretere *Inhaltsmodelle* lassen sich etwa mit Hilfe anderer Technologien zur Definition von *Inhaltsmodellen* aufstellen.²⁰⁷ Pauschal lässt sich sagen, je spezifischer die Verwendung der Daten ist, desto konkreter sind die Vorgaben zur Auszeichnung.

Als Beispiele für *XML*-Anwendungen können etwa die Guidelines der *Text Encoding Initiative*²⁰⁸ oder die Vorschläge zur *XML*-Strukturierung der *Dublin Core Metadata Initiative*,²⁰⁹ auf die ich im Folgenden noch eingehen werde, angeführt werden. Andere Anwendungen auf dem Gebiet von Texten sind beispielsweise das Open eBook Format (OeBF)²¹⁰ oder DocBook.²¹¹ Weitere bekannte *XML*-Anwendungen sind *XHTML*, eine *HTML*

206 Vgl. 3.1.4 Extensible Markup Language, S. 77.

207 Vgl. 3.2.4 XML-Inhaltsmodelle, S. 87.

208 Vgl. 3.3 Die Text Encoding Initiative, S. 94.

209 Vgl. Die Dublin Core Metadata Initiative, S. 133.

210 <http://www.idpf.org/> (18.05.2008).

211 <http://www.docbook.org> (18.05.2008).

Version, die auf der Grundlage von *XML* definiert wurde, die zur Kodierung mathematischer Formeln entwickelte Anwendung *Mathematical Markup Language* (MathML) sowie das für den Bereich grafischer Darstellung entwickelte Format *Scalable Vector Graphics* (SVG).

3.2.7 XML-Technologien

Mit der *XML*-Auszeichnung bzw. Kodierung von Texten oder Daten sind keinerlei Funktionen verbunden, *XML* ist nicht verfahrensorientiert, sondern dient lediglich der Strukturierung von Informationen. Um in einem *XML*-Dokument standardisierte Funktionen wie beispielsweise eine formatierte Ausgabe der Daten oder eine Verweisstruktur mittels *Hyperlinks*, wie sie in elektronischen Medien üblich sind, zu realisieren, benötigt man ergänzende Technologien. Die ergänzenden *XML*-Technologien lassen sich ihrem Zweck entsprechend in drei Gruppen aufteilen: Technologien zur Definition von *Inhaltsmodellen*, Technologien zur Formatierung und Technologien zur Adressierung. Technologien zur Definition von *Inhaltsmodellen* habe ich bereits im Abschnitt 3.2.4 *XML-Inhaltsmodelle* (S. 87) angesprochen. Auf Technologien zur Formatierung und Technologien zur Adressierung gehe ich im Folgenden kurz ein.

Technologien zur Formatierung

Eine der häufigsten Aufgaben beim Umgang mit elektronischen Texten ist die formatierte Darstellung bzw. Ausgabe der Daten auf dem Bildschirm, einem Drucker oder einer Druckmaschine. Konzepte eines »barrierefreien Internet« berücksichtigen darüber hinaus weitergehende Ausgabemöglichkeiten, beispielsweise die Sprachausgabe von Texten für Sehbehinderte. Zur Darstellungssteuerung haben sich zwei Technologien durchgesetzt: *Cascading Stylesheets* (CSS)²¹² und *Extensible Stylesheet Language*.²¹³

CSS ist eine ältere Technologie und wurde im Wesentlichen für die Formatierung von Dokumenten im *World Wide Web* entwickelt. Sie wird oft in Verbindung mit *HTML* verwendet, kann aber auch auf *XML*-Dokumente angewendet werden. Die Definition von CSS-Formaten erfolgt mit Hilfe einer eigenen Syntax. Mit Hilfe von *Cascading Stylesheets* können Inhalte für verschiedene Medien aufbereitet werden, etwa für die Ausgabe auf Bildschirm oder Drucker, aber auch die Steuerung einer Sprachausgabe ist möglich. Eine Einschränkung der CSS-Technologie besteht darin, dass sich die Formatierung immer auf die Inhalte aller Instanzen eines Elements bezieht. Die Formatierung einer bestimmten Instanz eines Elements oder einzelner Teile des Inhalts eines Elements ist nur möglich, wenn Formatanweisungen in das Dokument selber eingefügt werden. Daneben kann mit CSS kein Einfluss auf die Dokumentstruktur genommen werden, d. h. der Text kann nur in der Abfolge dargestellt werden, wie sie die Ausgangsdaten vorgeben.

XSL wurde zur Formatierung bzw. Ausgabesteuerung von *XML*-Dokumenten entwickelt. Die Möglichkeiten von *XSL* sind deutlich mächtiger und komplexer als die von CSS. Die Formatierung eines Dokuments kann spezifischer gesteuert werden, bis hin zur Formatierung einzelner Zeichen. Daneben kann auch die komplette Dokumentstruktur umgestellt werden. Um diese Möglichkeiten zu realisieren, arbeitet *XSL* anders als CSS. Während bei der CSS-gesteuerten Ausgabe das ursprüngliche Dokument nicht verändert wird, wird beim Arbeiten mit *XSL* ein neues Dokument generiert. Dazu besteht *XSL* aus zwei Teilen: Den *XSL-Formatting Objects* (*XSL-FO*), mit deren Hilfe ein Template erstellt wird, das vorgibt, wie die Formatierung aussehen soll, und den Anweisungen der *XSL Transformation* (*XSLT*), mit deren Hilfe festgelegt wird, auf welche Teile des Dokuments welche Formatierungsanweisungen des Templates angewendet werden.

Um bei der Transformation festlegen zu können, welche Teile eines Dokuments bei der Ausgabe wohin kommen und wie sie formatiert werden sollen, ist es notwendig, diese Teile bezeichnen, sie adressieren zu können.

212 <http://www.w3.org/Style/CSS/> (18.05.2008).

213 <http://www.w3.org/Style/XSL/> (18.05.2008).

XSLT greift zur Adressierung bzw. Auswahl der zu formatierenden Elemente und Inhalte auf *XML Path Language (XPath)*²¹⁴ zurück. Die XSL-Technologie erlaubt es damit, aus einem XML-Dokument beliebige Darstellungen für unterschiedliche Ausgabemedien zu generieren.

Technologien zur Adressierung

Technologien zur Adressierung werden benötigt, um innerhalb eines XML-Dokuments bestimmte Elemente oder deren Inhalte zu bezeichnen. Dies kann notwendig sein, um die Formatierung einer Textstelle oder eines Textbereiches festlegen zu können oder um auf eine bestimmte Stelle in einem Text verweisen zu können. Umgekehrt kann es notwendig sein, abfragen zu können, was an einer bestimmten Stelle des Dokuments steht. Die hierzu entwickelten Technologien sind *XPath*,²¹⁵ *XML Pointer Language (XPointer)*²¹⁶ und *XML Linking Language (XLink)*.²¹⁷

Die grundlegende Technologie ist dabei *XPath*. Die *XPath*-Technologie baut darauf auf, dass sich die Struktur eines XML-Dokuments als Baum darstellen lässt.²¹⁸ Alle möglichen Bestandteile eines XML-Dokuments – etwa Elemente, *Attribute* und Zeichendaten – werden als Knoten (nodes) betrachtet. Elemente, die andere Elemente enthalten, bilden Verzweigungen. Durch die Angabe eines Pfades, der den Verzweigungen folgend zu dem oder den gewünschten Knoten führt, können alle Knoten adressiert werden. Daneben stehen noch verschiedene Operatoren zur Verfügung, mit denen Elemente nach unterschiedlichen Aspekten ausgewählt werden können. Kriterien können beispielsweise *Attribute* oder Attributwerte von Elementen *s* (jedes Element *x*, das ein Attribut *a* mit dem Wert »123« hat) oder auch Elementinhalte sein (jedes Element *x*, das mehr als zehn und weniger als 20 untergeordnete Elemente *y* hat). Um innerhalb von Zeichendaten einzelne Bereiche, etwa einzelne Sätze oder Wörter adressieren zu können, stehen darüber hinaus sogenannte String-Operationen zur Verfügung.

Auf die Syntax von *XPath* gehe ich nicht weiter ein. Ein kurzes Beispiel macht das Konzept von *XPath* deutlich. Zur Veranschaulichung ziehe ich das oben angeführte Beispiel 4 heran, das hier ein eigenständiges XML-Dokument sein soll.

```
<Äußerung>
  <Sprecher>Danton</Sprecher>
  <Text>Spazieren, mein Junge, spazieren.
    <Regieanweisung>Er geht.</Regieanweisung>
  </Text>
</Äußerung>
```

Zeile aus DANTONS TOD semantisch ausgezeichnet

In dem Beispiel soll der Text der Regieanweisung »Er geht.« adressiert werden. Der entsprechende Adresspfad könnte folgendermaßen aussehen:

```
/Äußerung/Text[string (Regieanweisung)]
```

Zuerst wird der übergeordnete Knoten »Äußerung« ausgewählt, dann der untergeordnete Knoten »Text«. Unterhalb von »Text« wird zuletzt der Zeicheninhalt (string) des Knotens »Regieanweisung« als Ziel angegeben. Der Pfad wird deutlich, wenn man ihn in die Baumdarstellung des Beispiels einzeichnet (gepunktete Linie).

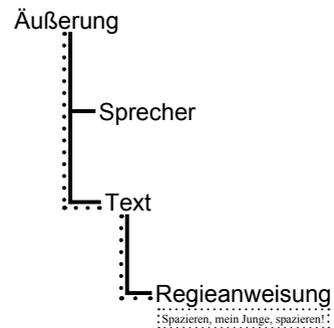


Abbildung 25. Veranschaulichung der XPath-Adressierung mit Hilfe eines Strukturbaums

214 Siehe in den Ausführungen zum Kafka Beispiel 3.3.4 Beispiel aus Franz Kafka: Der Process, S. 97.

215 <http://www.w3.org/TR/xpath> (18.05.2008) und <http://www.w3.org/TR/xpath20/> (18.05.2008).

216 <http://www.w3.org/XML/Linking> (18.05.2008).

217 <http://www.w3.org/XML/Query/> (18.05.2008).

218 Vgl. Abbildung 22. Darstellung der XML-Struktur als Strukturbaum, S. 87.

XPointer stellt im Wesentlichen eine Erweiterung von *XPath* dar, die es im Zusammenhang mit *XLink* erlaubt, auch dokumentübergreifende Verweise zu formulieren. D. h. mit Hilfe von *XPointer* und *XLink* kann theoretisch auf jedes einzelne Zeichen in allen *XML*-Dokumenten im *World Wide Web* verwiesen werden.

3.3 Die Text Encoding Initiative

Um den Umgang mit unterschiedlichen elektronischen Texten zu erleichtern, ist es wünschenswert, einheitliche Standards zur Auszeichnung von geisteswissenschaftlichen Texten zu etablieren. Zum einen müssen nicht für jeden Text neue Auszeichnungskonzepte entwickelt werden; zum anderen kann sich ein Nutzer schneller in einem fremden elektronischen Text zurechtfinden, wenn ihm die zugrundeliegenden Strukturen und Elemente bekannt sind. Eine standardisierte Datenstruktur erlaubt es darüber hinaus, Technologien und Programme, die für einen Text entwickelt wurden, auch für andere Texte zu nutzen, etwa *Stylesheets* oder *Retrieval*-Programme.

Die im Bereich der Literaturwissenschaft, Linguistik und Geschichte meistdiskutierte *XML*-Anwendung sind die von der *Text Encoding Initiative* entwickelten *GUIDELINES FOR ELECTRONIC TEXT ENCODING AND INTERCHANGE*. Die Initiative ging aus dem Zusammenschluss dreier wissenschaftlicher Gesellschaften, der *Association for Computers and the Humanities*, der *Association for Computational Linguistics* und der *Association for Literary and Linguistic Computing* hervor und wurde auf breiter Ebene unterstützt, etwa von der amerikanischen *National Endowment for the Humanities* oder der Europäischen Union. Sie wurde 1987 gegründet. Im Jahr 2000 ging aus ihr eine Arbeitsgemeinschaft hervor, die seitdem als gemeinnützige Gesellschaft die *GUIDELINES FOR ELECTRONIC TEXT ENCODING AND INTERCHANGE* betreut und weiterentwickelt.

Die *TEI*-Arbeitsgemeinschaft versucht auf internationaler Ebene allgemeine, für den wissenschaftlichen Gebrauch verwendbare Richtlinien zur Textauszeichnung zu etablieren. Ursprünglich wurden die Richtlinien zur Textauszeichnung, bis zur Version *TEI P3*, auf Basis von *SGML* entwickelt. Die darauf verabschiedeten Richtlinien, *TEI P4*, beruhen auf *XML*, das zugehörige *Inhaltsmodell* wurde in Form einer *DTD* angegeben. Die neueste Version, *TEI P5*, wurde Ende 2007 verabschiedet. Die wichtigste Neuerung war dabei die Umstellung des *Inhaltsmodells* auf *RELAX-NG*.

Neben Weiterentwicklung und Pflege der *TEI*-Richtlinien und der Formulierung geeigneter *Inhaltsmodelle* ist ein weiterer Schwerpunkt der *Text Encoding Initiative* die Entwicklung, Förderung und Verbreitung von Programmen und Werkzeugen zum Umgang mit *TEI*-kodierte Texten. Dazu gehören besonders Hilfsmittel zur Anpassung des *TEI*-Schemas an spezielle Bedürfnisse bei der Textauszeichnung sowie *Stylesheets* zur formatierten Darstellung *TEI*-kodierter Texte und Werkzeuge zur Publikation. Die Möglichkeit, beim Umgang mit elektronischen Texten auf vorhandene Programme und Werkzeuge zurückgreifen zu können oder auf bestehenden Konzepten aufbauen zu können, bedeutet eine nicht unerhebliche Erleichterung im Umgang mit elektronischen Texten, insbesondere wenn man berücksichtigt, dass der Umgang mit Computertechnologien oder die Entwicklung von Computerprogrammen nur beschränkt in das Aufgabenfeld der Geisteswissenschaften fällt.

Die *TEI*-Richtlinien stellen derzeit einen quasi-Standard bei der geisteswissenschaftlichen Textauszeichnung dar. Auch wenn es von deutscher Seite noch keine offizielle Beteiligung am *TEI*-Consortium gibt, wird eine Orientierung an den hier entwickelten Richtlinien von verschiedenen Stellen gefordert, empfohlen oder bereits praktiziert. Anzuführen sind hier etwa die »Digitalisierung des DEUTSCHEN WÖRTERBUCHS VON JACOB UND WILHELM GRIMM; dessen Publikation auf CD-ROM und Bereitstellung im Internet«²¹⁹ oder *DER JUNGE GOETHE IN SEINER WELT*.²²⁰ Die Deutsche Forschungsgemeinschaft (DFG) hat in ihrem Anforderungsprofil für

219 <http://germazope.uni-trier.de/Projects/DWB> (18.05.2008).

220 <http://www.jgoethe.uni-muenchen.de/> (18.05.2008).

das 1997 gestartete Förderungsprogramm »Retrospektive Digitalisierung von Bibliotheksbeständen«²²¹ bei der Textauszeichnung eine Orientierung an *TEI* empfohlen.

3.3.1 *TEI*-Richtlinien

Wie ein Text für elektronische Medien ausgezeichnet wird, hängt im Wesentlichen von Art und Aufbau des Textes selber und dem angestrebten Verwendungszweck des Textes ab. Sowohl die zu strukturierenden Texte, als auch die vorgesehene Verwendung können sehr unterschiedlich sein. Die *TEI*-Richtlinien tragen diesem Umstand Rechnung, indem sie keine Regeln zur Textauszeichnung festlegen, sondern lediglich sehr weit gefasste Empfehlungen geben.

Dokumentstruktur

Die *TEI*-Richtlinien sehen zunächst die Einteilung eines *TEI*-Dokuments in einen Dokumentkopf, den »*TEI*-Header«, und den Text vor. Der *TEI*-Header ist dabei deskriptiven Metadaten²²² vorbehalten. Auf die Funktion und Struktur des *TEI*-Headers gehe ich im Kapitel 3.4 Metadaten noch ein.²²³

Für den Text selber sehen die *TEI*-Richtlinien eine Gliederung in die Bereiche »front«, »body« und »back« vor. Das Element »front« enthält dabei alle dem eigentlichen Text vorangestellten Informationen wie den Text der Titelseite, Vorwörter oder Widmungen etc. Der Bereich des »body« enthält den eigentlichen Text. Das Element »back« beinhaltet schließlich alle dem eigentlichen Text nachgestellten Informationen wie etwa Anhänge zum Haupttext.

Die Grundstruktur eines Textes, der nach *TEI*-Richtlinien ausgezeichnet wird, sieht folgendermaßen aus:

```
<TEI>
  <teiHeader> Hier stehen Metadaten zum Text </teiHeader>
  <text>
    <front>
      Hier stehen dem Text vorangestellte Informationen
    </front>
    <body>
      Hier steht der eigentliche Text
    </body>
    <back>
      Hier stehen dem Text nachgestellte Informationen
    </back>
  </text>
</TEI>
```

Die Struktur ist sehr allgemein gehalten. Die meisten Texte dürften in diese Struktur einzuordnen sein. Sollte ein Text aus mehreren einzelnen Bereichen, die jeweils »front«- und/oder »back«-Elemente enthalten, bestehen, kann jeder dieser Bereiche mit Hilfe des Elements »group« noch zu einer Untergruppe des eigentlichen Gesamttextes zusammengefasst werden.

Textauszeichnung

Die genaue Auszeichnung kann in Abhängigkeit von der Art des Textes und dem vorgesehenen Verwendungszweck erfolgen. Um die Auszeichnung unterschiedlichster Textgrundlagen zu unterstützen, gibt *TEI* einen Satz von Elementen vor, die in allen auszuzeichnenden Texten vorkommen können (Elements available in all *TEI*-Documents) und bereits zur Strukturierung der meisten Texte ausreichen. Diese können, je nach Bedarf, durch

²²¹ Altrichter, Helmut: Retrodigitalisierung in Deutschland – Versuch einer Zwischenbilanz.

<http://www.bsb-muenchen.de/mdz/forum/altrichter/> (18.05.2008).

²²² Vgl. Deskriptive Metadaten, S. 129.

²²³ Vgl. 3.4 Metadaten, S. 126.

Elemente für spezielle Textsorten oder Anforderungen ergänzt werden. Spezielle Gruppen von Elementen stehen etwa für Lyrik, Dramatik, Wörterbücher und Lexika oder die Transkription von mündlicher Sprache oder handschriftlichen Quellen zur Verfügung.

Bei der Handhabung der einzelnen Elemente und ihrer Ergänzung durch *Attribute* lassen die *TEI*-Richtlinien einen weiten Gestaltungsspielraum. Die genaue Textauszeichnung und die Einhaltung einer einheitlichen Struktur obliegen somit der Kontrolle der für die Textauszeichnung Verantwortlichen. Beispiele für verschiedene Möglichkeiten diskutiere ich weiter unten.²²⁴

Grundsätzlich müssen bei der *XML*-Kodierung nach *TEI*-Richtlinien ähnliche Überlegungen wie bei der Erstellung einer Buchedition getroffen werden. Auch hier ist es notwendig, das Ziel der Edition zu formulieren und die geeigneten Mittel, etwa die Art der Darstellung, die Ausgestaltung des Apparates etc., zu wählen. Im Fall elektronischer Editionen ist es jedoch notwendig, dass der Herausgeber neben den typografischen und herstellerischen Mitteln, die ihm für eine Buchedition zur Verfügung stehen, auch die technischen Möglichkeiten und Anforderungen elektronischer Medien kennt und berücksichtigt. Dies betrifft sowohl Techniken zur Darstellung wie *XSL* als auch die Anforderungen von Datenbanken und Internettechnologien. Je nach Material und Verwendungszweck wird man sich im Rahmen der *TEI*-Richtlinien für eine bestimmte Art der Textauszeichnung entscheiden.

Besonderheiten oder Konventionen, die im Rahmen einer Edition gemacht werden und auf die bei einer Buchausgabe in der Einleitung, den editorischen Bemerkungen oder speziellen Verzeichnissen für verwendete Abkürzungen, Siglen, oder Zeichen sowie Schriftarten- und -auszeichnungen angegeben werden, können auch im Falle von *TEI*-kodierte Texten notwendig sein. Entsprechende Informationen werden im *TEI*-Header angegeben.²²⁵

3.3.2 *TEI*-Inhaltsmodell

Um bei aller Flexibilität der Textauszeichnung nach den *TEI*-Richtlinien noch kontrollieren zu können, ob ein Text *TEI*-konform ausgezeichnet wurde, wurde für die *TEI*-Richtlinien bis P₄ jeweils eine *DTD* entwickelt, die modular aufgebaut war, um dem Umstand möglichst großer Flexibilität Rechnung zu tragen. Die einzelnen *DTD*-Module konnten in Abhängigkeit von den benutzten Element-Sets eingebunden werden. Eine nicht geringe Schwierigkeit bei der Textauszeichnung nach den *TEI*-Richtlinien bestand darin, dass man sich eine geeignete *DTD* zusammenstellen musste, wollte man ein gültiges *TEI*-Dokument erstellen.²²⁶

Das Inhaltsmodell der aktuellen *TEI*-Richtlinien P₅ wird mit Hilfe von *RELAX-NG* definiert. Der Vorteil dabei ist, dass *RELAX-NG* es ermöglicht, unkomplizierter ein den Erfordernissen der geisteswissenschaftlichen Textauszeichnung gemäßes Schema zu definieren. Nachteilig ist dabei, dass *RELAX-NG* bei weitem nicht so etabliert ist wie die *DTD*-Technologie. Entsprechend kennen sich weniger Menschen mit *RELAX-NG* aus, und es steht keine so umfangreiche Auswahl an Software zur Verfügung, die *XML*-Dokumente gegen ein *RELAX-NG*-Schema validieren kann.

Auch das *RELAX-NG*-Schema der P₅-Richtlinien ist modular aufgebaut, so dass man sich aus den einzelnen Modulen ein geeignetes Schema zusammenstellen kann.

²²⁴ Vgl. 3.3.4 Beispiel aus Franz Kafka: *DER PROCESS*, S. 97 und 3.3.5 Beispiel aus Georg Büchner: *DANTONS TOD*, S. 108.

²²⁵ Vgl. 3.4.4 Kodierung von Metadaten. S. 131.

²²⁶ Auf die Verwendung der *DTD* für die Richtlinien *TEI* P₃ geht ausführlich ein: Jannidis, Fotis: *TEI in der Praxis. Arbeitsbericht*: <http://computerphilologie.tu-darmstadt.de/praxis/teipraxis.html> (18.05.2008).

3.3.3 Textkodierung nach TEI-Richtlinien

Im deutschsprachigen Raum finden sich wenige Projekte, bei denen eine Textauszeichnung nach *TEI*-Richtlinien konsequent umgesetzt wurde. Die umfangreichste Erfahrung auf diesem Gebiet liegt bei digitalen Wörterbüchern vor. Bei der Retrodigitalisierung mittelhochdeutscher Wörterbücher wurde weitgehend nach *TEI*-Richtlinien gearbeitet.²²⁷ Auch im Rahmen des Projekts *DAS DEUTSCHE WÖRTERBUCH VON JACOB GRIMM UND WILHELM GRIMM AUF CD-ROM UND IM INTERNET* hat man sich an *TEI* orientiert.²²⁸ Wörterbücher scheinen aufgrund ihrer Struktur eine hohe Affinität zu elektronischen Medien aufzuweisen. Ihre Struktur aus Stichwörtern und Verweisen bietet sich etwa für *Hypertext*-Konzepte an. Daneben sind sie in sich meist klar gegliedert und kommen den Strukturen von Datensätzen nahe.

Literarische Texte, die in einer elektronischen Form nach *TEI* ausgezeichnet wurden, gibt es bislang kaum. Anzuführen ist hier etwa die *TEI*-Version der Hybrid-Edition *DER JUNGE GOETHE IN SEINER ZEIT*.²²⁹ Leider sind die Informationen hierzu spärlich.²³⁰

Im Folgenden stelle ich zwei Beispiele einer *TEI*-konformen Kodierung elektronischer Texte vor. Dabei gehe ich von etablierten historisch-kritischen Buchausgaben aus: Von der *FRANKFURTER KAFKA AUSGABE*²³¹ (FKA) und der *MARBURGER BÜCHNER AUSGABE*²³² (MBA).

Der Tenor dieser Arbeit zielt darauf ab, für ein eigenständiges Konzept elektronischer Editionen zu plädieren, aus denen als eine mögliche Darstellungsform eine Buchausgabe generiert werden könnte. Ein eigenständiges, komplett elektronisch angelegtes Editionsprojekt würde den Rahmen dieser Arbeit jedoch sprengen. Bei den angeführten Beispielen gehe ich deshalb von etablierten Buchausgaben aus. In einem ersten Schritt werde ich das Material vorstellen und die im Buch enthaltenen Informationen explizit aufführen, d. h. die typografische Kodierung entschlüsseln. In einem zweiten Schritt werde ich jeweils einen Vorschlag vorstellen, wie ein entsprechendes elektronisches Kodierungsmodell, das den *TEI*-Richtlinien entspricht, aussehen könnte. Exemplarisch greife ich dabei Punkte auf, die Möglichkeiten und Grenzen einer *TEI*-konformen Textauszeichnung aufzeigen. Es handelt sich bei den von mir vorgestellten Textausschnitten um Beispiele, die lediglich verdeutlichen sollen, welche Überlegungen bei der elektronischen Kodierung von Editionen zu berücksichtigen sind, keinesfalls um komplette Konzepte.

3.3.4 Beispiel aus Franz Kafka: DER PROCESS

Erklärtes Ziel der editorischen Bemühungen der Frankfurter Kafka Ausgabe ist es, aus dem Nachlass Kafkas einen »endgültigen« Text zu konstruieren oder Kafka »zu Ende zu dichten«:

227 Burch, Thomas und Fournier, Johannes: Zur Anwendung der *TEI*-Richtlinien auf die Retrodigitalisierung mittelhochdeutscher Wörterbücher. In: Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. Hg. von Ingrid Lemberg u. a. Tübingen: Niemeyer, 2001. (Lexicographica: Series maior 107.). S. 133-153.

228 Hildenbrandt, Vera und Schares, Thomas: Das Grimmsche Wörterbuch geht ins 21. Jahrhundert: Präsentation eines Prototyps des digitalen Deutschen Wörterbuchs von Jacob und Wilhelm Grimm. <http://www.bsb-muenchen.de/mdz/forum/hildenbrandt/index.htm> (18.05.2008); Burch, Thomas u. a.: Ein »Hausbuch« für alle? – Das Deutsche Wörterbuch von Jacob und Wilhelm Grimm auf CD-ROM und im Internet. In: Jahrbuch für Computerphilologie 2 (2000). S. 11-34; Christmann, Ruth u. a.: Ein »heiligthum der sprache« digitalisiert: Das Deutsche Wörterbuch von Jacob und Wilhelm Grimm auf CD-ROM und im Internet. In: TUSTEP educa. Actas de Congreso del International TUSTEP User Group. Penaranda de Duero (Burgos) Octubre 1999. Hg. von Nicolás Benito Castrillo und Peter Stahl. Burgos: Univ. 2001. S. 13-37; Christmann, Ruth: Books into Bytes: Jacob and Wilhelm Grimm's Deutsches Wörterbuch on CD-ROM and on the Internet. *Literary and Linguistic Computing* 1 (2001). S. 121-133.

229 Goethe, *Der junge Goethe in seiner Zeit*.

230 Der Vollständigkeit halber erwähnen möchte ich noch das *PROJEKT LORELEY*, das 1999 an der Humboldt Universität Berlin begonnen wurde und einige nach *TEI P3 (Lite)* ausgezeichnete Gedichte präsentiert.

231 Kafka, Franz: *Historisch-kritische Ausgabe sämtlicher Handschriften, Drucke und Typoskripte*.

232 Büchner, Georg: *Sämtliche Werke und Schriften. Historisch-kritische Ausgabe mit Quellendokumentation und Kommentar*. Hg. von Burghard Dedner und Thomas Michael Mayer. 14 Bde. Darmstadt: Wissenschaftliche Buchgesellschaft, 2000 ff.

Zu einem definitiven Text der Handschrift (im konventionellen Sinne) gelangt die FKA, will sie wiedergeben, was überliefert ist, nicht.²³³

Was im Zusammenhang des »Verschollenen«, des »Processes« und des »Schlusses« überliefert ist, sind mehr oder weniger weit gediehene Entwürfe, deren Fragmentcharakter [...] nicht akzidentell ist, sondern mit der Sache von Kafkas Schreiben innigst zusammenhängt. Jeder noch so gut gemeinte Versuch der Philologie, hieraus einen Text herzustellen, der über das Stadium hinausgeht, in dem Kafka sein Schreiben abbrach [...], ist gegenüber diesem Schreiben unernst.²³⁴

Entsprechend bietet die FKA im Wesentlichen die faksimilierten Manuskripte und Typoskripte aus dem Nachlass Kafkas und deren differenzierte Umschrift. Um dem fragmentarischen Charakter der überlieferten Schriften Rechnung zu tragen, verzichtet die Ausgabe sogar darauf, die überlieferten Manuskripte und Typoskripte in einem geschlossenen Buch-Band zu veröffentlichen und präsentiert das Material in Form der von Kafka zusammengestellten Konvolute, die jeweils ein eigenes Heft innerhalb eines »Bandes« bilden. Der »Prozess-Band« der FKA besteht aus einem Schuber mit 17 einzelnen Heften, je ein Heft für die 16 Konvolute zuzüglich des »Franz Kafka-Heftes 1«, das Erläuterungen zum Text und zur Edition enthält. Im Anhang eines jeden Konvoluts wird die Herkunft der einzelnen, zum Konvolut gehörenden Blätter nachgewiesen, die meist aus den von Kafka häufig benutzten Quartheften stammen, die zur Zusammenstellung der einzelnen Konvolute auseinandergenommen wurden.

Den einzelnen Schubern der gedruckten FKA liegen CD-ROMs bei, die das Material der Buchausgabe im *PDF* enthalten.

PDF-Edition der FKA

Den einzelnen Bänden der FKA liegt jeweils eine CD-ROM bei, die das als Buch gedruckte Material im *PDF* enthält. Es handelt sich somit, wie oben erwähnt, um eine buchnahe elektronische Edition. Der Begleittext der Edition bemerkt dazu:

Ihr Hauptzweck ist die *Information*; sie dient nicht, was immer derzeit metaphorisch gängige Münze sein mag, dem *Lesen*. Als spezifisches Informationsmedium ist sie ein Werkzeug zu Zwecken der Recherche, nicht mehr, aber auch nicht weniger. Sie ermöglicht mit ihrem vollständigen Zeichenindex das Suchen nach Wortfeldern und Begriffen – mit dem Ziel, die Herstellung von Kontur zu erleichtern, nicht: sie zu negieren und als superflu erscheinen zu lassen. Womit sie, richtig benutzt, immer wieder auf die Handschrift und die typographische Übersetzung der Hefte zurückverweist.²³⁵

Damit charakterisiert Reuß die elektronischen Texte der CD-ROM treffend: Sie verweisen auf die »typographische Übersetzung« der Handschrift zurück. Konsequenterweise wird die CD-ROM nicht als elektronische Edition vertrieben, sondern nur im Zusammenhang mit den Büchern.

Faksimile der Handschrift

Bei dem abgebildeten Faksimile der Handschrift handelt es sich um die erste Seite des Konvoluts, das Kafka mit »Der Onkel Leni« betitelt hat. Die FKA dokumentiert auch die Herkunft der Blätter: Das Blatt stammt aus einem der von Kafka für seine Notizen und Tagebücher verwendeten Quarthefte, dem verschollenen bzw. aufgelösten Processheft 3.

233 Reuß, Roland: Lesen, was gestrichen wurde. Für eine historisch-kritische Kafka-Ausgabe. In: Franz Kafka: Historisch-kritische Ausgabe sämtlicher Handschriften und Typoskripte. Hg. von Roland Reuß. Einleitungsband. Basel, Frankfurt a. M., 1995. S. 19.

234 Reuß (1995), S. 21.

235 Roland Reuß: Zur kritischen Edition von »Der Prozess«. In: Franz Kafka: Historisch-kritische Ausgabe sämtlicher Handschriften und Typoskripte. Hg. von Roland Reuß. Bd. 1: Der Prozess, Buch: Franz Kafka-Hefte 1. Basel; Frankfurt am Main: Stroemfeld, 1997. S. 24.

<1> Eines Nachmittags – K. war gerade vor dem Postabschluss sehr 134
 beschäftigt – drängte sich zwischen zwei Diener'n, die Schriftstücke
 herein
 trugen ~~Onkel Karl~~ K.'s Onkel Karl, ein kleiner Grund-
 besitzer vom Lande, ins Zimmer. K. erschrak bei dem 5
 Anblick weniger, als er schon vor längerer Zeit bei der
 Vorstellung vom Kommen des Onkels ersch[ak]locken] Der
 Onkel musste kommen, dass stand bei K. schon etwa einen
 Monat lang fest. Schon damals hatte er ihn zu sehen 10
 ein wenig gebückt
 geglaubt, wie er er den eingedrückten Panamahut
 in der Linken die Rechte schon von weitem ihm
 entgegenstreckte und mit rücksichtsloser Eile über 15
 rücksichtslos alles umstossend, was ihm im Wege war
 den Schreibtisch hin reichte, auf dem bei dieser
 rechts und links befand
 Gelegenheit etwas umfief. oder gar Der Onkel war 20
 immer in Eile, denn er war von dem Gedanken
 immer nur
 verfolgt bei seinem eintägigen Aufenthalt in der
 Hauptstadt müsse er alles erledigen können, was er 25
 dürfe, auch
 sich vorgenommen hatte und überdies kein gelegentlich
 sich anbietendes Gespräch oder Geschäft oder Vergnügen
 sich entgehen lassen. Dabei musste ihm aber K., der
 ihm als seinem gewesenen Vormund besonders verpflichtet war,
 in allem möglichen behilflich sein und ihn ausserdem 30
 bei sich übernachten lassen. „Das Gespenst vom Lande“
 pflegte er ihn zu nennen.
 Gleich nach der Begrüssung – sich in das Fauteuil
 zu setzen, wozu ihn K. einlud, hatte er keine Zeit. Er bat 35
 unter vier
 er K. um ein kurzes Gespräch zwischen ~~zwei~~ [U]Augen. „Es

6
 1 134] <blau]stift>
 35 vier] <violette Tinte>
 36 ~~zwei~~] <Durchstreichung mit violetter Tinte>

Abbildung 27. Differenzierte Umschrift der ersten Manuskriptseite des Kapitels »Der Onkel Leni« Kafka, Franz: Historisch-kritische Ausgabe sämtlicher Handschriften, Drucke und Typoskripte. Bd. 1: Der Process. Buch: Der Onkel Leni. S. 6

Differenzierte Umschrift

Die differenzierte Umschrift enthält im Wesentlichen die Informationen, die als Grundlage für einen elektronischen Text dienen können. Die typografische Kodierung der Befunde der Handschrift, wie sie der Text bietet, muss auch in diesem Fall in eine elektronische, semantische Kodierung umgesetzt werden. Der Buchvorlage sind folgende Informationen zu entnehmen (die angegebenen Zeilennummern beziehen sich auf die differenzierte Umschrift in Abbildung 27):

- die Seite trägt die von Max Brod eingefügte Paginierung Nummer 134
- in Zeile 2 ergänzte Kafka das Wort »Diener« um ein »n« zu »Dienern«
- in Zeile 4 ergänzte er das Wort »trugen« durch ein über der Zeile eingefügtes »herein« zu »hereintrugen«
- in der gleichen Zeile wurden die Wörter »Onkel Karl« gestrichen
- in der Zeile 8 wurde das Wort »erschrak« durch Überschreibung des »ak« mit einem »o« und der Einfügung des Wortes »war« zu »erschrocken war« korrigiert
- Zeile 22 wurde durch die Einfügung aus Zeile 21 »ein wenig gebückt« ergänzt
- in Zeile 15 wurde das »sie« ergänzt, indem es über der Zeile als Einfügung notiert wurde
- das Ende der Zeile 17 wurde gestrichen und durch den darüber notierten Text »rücksichtslos alles umstossend, was ihm im Wege war« ersetzt, wobei das Wort »rücksichtslos« wieder gestrichen wurde
- in Zeile 19 wurde hinter »Gelegenheit« »rechts und links« darübergeschrieben ergänzt, der Text »Gelegenheit etwas umfiel« gestrichen und dahinter ein Punkt ergänzt; das »war« am Ende der Zeile wurde gestrichen und ein darüber notiertes »befand« eingefügt
- in Zeile 21 wurde das Wort »unglücklichen« zwischen den Zeilen zwischen »dem« und »Gedanken« eingefügt
- auf die gleiche Weise wurden in Zeile 23 die Worte »immer nur« und in Zeile 26 die Worte »dürfe« und »auch« ergänzt
- in Zeile 28 wurde das »aber« gestrichen
- Zeile 34 wurde das Satzende »Zeit.« zu »Zeit – « korrigiert
- in Zeile 36 wurde das Wort »zwischen« gestrichen und das Wort »unter« darüber ergänzt; die Ziffer »4« wurde von Max Brod gestrichen und durch das darüber das Wort »vier« notiert.

Hier das Beispiel für eine mögliche XML-Kodierung des Textes

Textauszeichnung nach TEI

```
<?xml version="1.0" encoding="UTF 8"?> 1
<TEI>
  <teiHeader>
    <fileDesc>Beispiel zur Textauszeichnung nach TEI</fileDesc>
  </teiHeader> 5
  <text>
    <body>
      <!-- ... -->
    </body>
  </text>
  <pb ed="Der Onkel Leni" n="1"/>
  <milestone unit="Blattherkunft" ed="Prozessheft 3" n="1r"/> 10
  <p>
    <b n="1"/>Nachmittags – K. war gerade vor dem Postabschluss sehr
      <add hand="Max_Brod" place="inline">134</add>
      <note>Blaustift</note>
    <b n="2"/>beschäftigt – drängte sich zwischen zwei Diener<corr>n</corr>, 15
    die Schriftstücke
    <b n="4"/>
      <add n="3" place="supralinear" rend="addline">herein</add>trugen
```



```

      <add n="35" place="supralinear" hand="Max_Brod">vier</add>
      <note>violette Tinte</note>
Augen. „Es
<!-- ... -->
</p>
  </body>
  </text>
</TEI>

```

80

85

Beispiel 10. XML/TEI kodierter Text der ersten Manuskriptseite aus Franz Kafka: Der Onkel Leni

Alle im Beispiel verwendeten Auszeichnungen sind bereits im *TEI* »core modul« enthalten, das allgemeine, in allen *TEI*-Dokumenten zur Verfügung stehende Elemente beschreibt. Im Folgenden gehe ich auf die Einzelheiten der Kodierung des Beispiels ein. Die Zeilennummern beziehen sich auf die Zeilenzählung des *XML-/TEI*-kodierten Textes (Beispiel 10).

Zeile 1 enthält die bei jedem *XML*-Dokument erforderliche *XML*-Deklaration. Diese ist notwendig, da *TEI* eine *XML-Anwendung* ist, also auf den Auszeichnungsregeln von *XML* beruht.

In Zeile 2 markiert ein öffnender Tag den Beginn des Root-Elements, das einfach »*TEI*« heißt.

Die Zeilen 3 bis 5 enthalten den für dieses Beispiel verkürzten *TEI*-Header.²³⁶

Der eigentliche Text beginnt in Zeile 6. Da das Beispiel keine dem eigentlichen Text voran- oder nachgestellten Textteile enthält, können die Elemente »front« und »back« entfallen. Der Textkörper beginnt in Zeile 7 mit dem Tag »body«.

Der Kommentar in Zeile 8 soll lediglich darauf hinweisen, dass es sich bei dem folgenden Text um einen Ausschnitt handelt.

Die Kodierung der als Beispiel herangezogenen Manuskriptseite beginnt in Zeile 9:

```
<pb ed="Der Onkel Leni" n="1"/>
```

Das Element »pb« markiert einen Seitenumbruch (pagebreak). Die *TEI*-Richtlinien ordnen das Element den sogenannten Meilenstein-Elementen (milestone elements) zu. Meilenstein-Elemente sind zur Kennzeichnung nicht hierarchischer Strukturen vorgesehen und deshalb als leere Elemente deklariert. Für die hierarchische Struktur von Prosatexten relevante Gliederungsmerkmale sind meist Absätze, die nicht zwingend mit Seitenumbrüchen zusammenfallen. Oft gehen Absätze über Seitenumbrüche hinaus. Insofern hat die Gliederung eines Prosatextes in Absätze Vorrang bei der Textauszeichnung. Um Probleme mit sich überschneidenden Elementen zu vermeiden, ist die Verwendung eines leeren Elements in diesem Fall von Vorteil.²³⁷

Das Element »pb« enthält die *Attribute* »ed« und »n«. Das fakultative Attribut »ed«, Edition, kann dem »pb« Element beigefügt werden, um anzugeben, in welcher Textausgabe an dieser Stelle ein Seitenumbruch erfolgt. In diesem Fall bezieht sich der Seitenumbruch auf die Seiten des Textkonvoluts »Der Onkel Leni«. Das Attribut »n« kann jedem *TEI*-Element beigefügt und für jede beliebige Form der Zählung verwendet werden. Da es sich um die erste Seite des Konvoluts handelt, habe ich hier die Nummer »1« vergeben.

Informationen wie diese können für die spätere Darstellung auf unterschiedliche Weise ausgewertet werden. Die *SGML*-Datei, die den elektronischen Editionen *KAFKAS WERKE IM WWW* bzw. *KAFKAS WERKE AUF CD-ROM*

²³⁶ Vgl. 3.3.1 *TEI*-Richtlinien, S. 95.

²³⁷ Vgl. Struktur, S. 83.

zugrunde liegt, enthält *Tags*, die Seitenumbrüche der Buchvorlage markieren.²³⁸ Für die Bildschirmdarstellung wird diese Information genutzt, um an den Stellen, an denen in der Buchvorlage ein Seitenumbruch erfolgt, den Textfluss zu unterbrechen und die Buchseitennummer sowie *Hyperlinks* zum Apparat einzufügen.²³⁹

Einen anderen Weg geht die Edition DEUTSCHE LITERATUR VON LESSING BIS KAFKA.²⁴⁰ Auch hier sind Informationen über die zugrundeliegende Ausgabe sowie die Seitenzahl des Textes der Vorlage in der *XML*-Datei kodiert worden. Diese Edition verwendet bei der Bildschirmdarstellung eigene, feste Seitenumbrüche, die nicht mit derjenigen der Vorlage übereinstimmen.

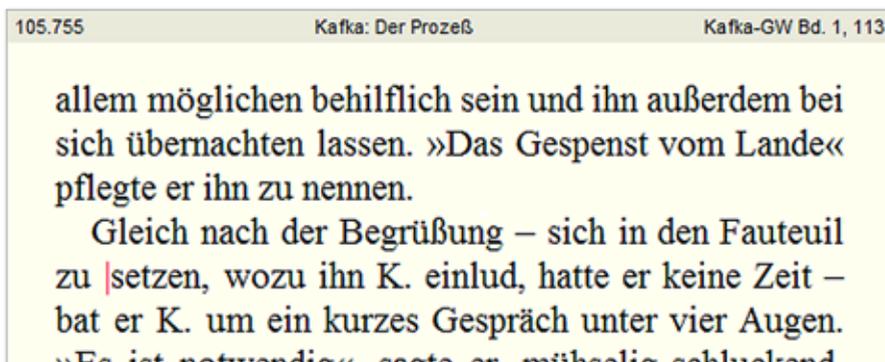


Abbildung 28. Ausschnitt einer Seite der CD-ROM DEUTSCHE LITERATUR VON LESSING BIS KAFKA

In der Kopfzeile jeder Seite der elektronischen Edition werden dabei mittig Autor und Werktitel angezeigt. Links in der Kopfzeile steht die Seitenzahl der elektronischen Edition (105755), rechts zuerst ein Sigel, das auf die zugrunde gelegte Buchausgabe verweist (Kafka-GW Bd. 1), sowie die Seitenzahl, auf der der Text in der Buchausgabe zu finden ist. Da die Seitenumbrüche der elektronischen Edition nicht mit der Buchausgabe übereinstimmen, wird an Stellen, an denen in der Buchvorlage ein Seitenumbruch erfolgt, eine dünne rote Linie angezeigt, um den Seitenumbruch der Vorlage zu markieren (in Abbildung 28 am Beginn der fünften Zeile zwischen »zu« und »setzen«). Die in der Kopfzeile angezeigte Seitenzahl richtet sich nach der Position des Mauszeigers. Fährt man mit der Maus über einen Seitenumbruch der Vorlage hinweg, ändert sich die Anzeige in der Kopfzeile. Die eingepflegten Konkordanzangaben werden auch in anderen Bereichen der Edition genutzt. Beispielsweise wird beim Kopieren von Text aus der elektronischen Edition automatisch ein Quellennachweis generiert und angefügt. Auch die Fundstellenlisten der Digitalen BIBLIOTHEK enthalten eine entsprechende Quellenangabe.²⁴¹

»Das Gespenst vom Lande« pflegte er ihn zu nennen.

[Kafka: Der Prozeß. Deutsche Literatur von Lessing bis Kafka, S. 105755 (vgl. Kafka-GW Bd. 1, S. 112)
<http://www.digitale-bibliothek.de/band1.htm>]

Beispiel 11. Aus der elektronischen Edition DEUTSCHE LITERATUR VON LESSING BIS KAFKA kopierte Textzeile mit automatisch generierter Quellenangabe

Da die Herkunft der Blätter innerhalb der einzelnen Konvolute, in denen das Werk Kafkas überliefert ist, unterschiedlich ist, habe ich in der Zeile 10 ein weiteres Meilenstein-Element, das »milestone«-Element, eingefügt.

```
<milestone unit="Blattherkunft" ed="Prozessheft 3" n="1r"/>
```

238 Vgl. Beispiel 2. Auszug aus dem SGML-Quellcode der elektronischen Editionen KAFKAS WERKE IN WWW und KAFKAS WERKE AUF CD-ROM. S. 75 f. Das entsprechende Element hat hier ebenfalls den Namen »pb«.

239 Vgl. Abbildung 4 Screenshot Kafkas Werke im WWW. Kritische Kafka-Ausgabe des S. Fischer Verlages bei Chadwyck-Healey. Chadwyck-Healey, 1999. S. 15.

240 Bertram (2000), S. 105775.

241 Vgl. Abbildung 8 Ausschnitt aus der Fundstellenliste der DIGITALEN BIBLIOTHEK des Verlags Directmedia Publishing. S. 28.

Das »milestone«-Element habe ich verwendet, um die Blattherkunft, die in der FKA im Anhang des Konvoluts dokumentiert wird, textnah anzugeben. Das Attribut »unit« dient zur Spezifikation des »milestone«-Elements. Hier soll gemäß den *TEI*-Richtlinien angegeben werden, um welche Art von Struktur es sich handelt. In diesem Fall gibt das »milestone«-Element die Blattherkunft an. Das »ed«-Attribut habe ich benutzt um anzugeben, dass das Blatt aus dem sogenannten Processheft 3 stammt. Das Zählattribut »n« gibt an, dass es sich um die Rückseite des ersten Blattes des Quartheftes, also »eins recto«, »1r« handelt.

Diese Information könnte in der Darstellung einer elektronischen Edition ähnlich verwendet werden wie die des Seitenumbruchs.

Der Text selber ist, wie es bei den meisten Prosatexten der Fall ist, in Absätze gegliedert. Jeder Absatz ist als Instanz des Absatz-Elements »p« ausgezeichnet. Entsprechende *Tags* markieren den Beginn und Ende der einzelnen Absätze. Der erste Absatz des Beispieltexes beginnt in Zeile 11 mit dem öffnenden »p«-Tag.

Die Zeilenzählung der Buchausgabe ist im *TEI*-kodierte Text mit Hilfe des Elements »lb«, linebreak (Zeilenumbruch, ab Zeile 12) eingepflegt: `<lb n="1"/>`. Auch das »lb«-Element gehört zu den Meilenstein-Elementen, da Zeilenumbrüche in Prosatexten in den meisten Fällen keine Gliederungsmerkmale des Textes sind. Dass im oben angeführten Beispieltext jede Textzeile auf einer eigenen Zeile beginnt, dient lediglich der übersichtlichen Darstellung. Leerzeichen, Tabulatoren und Zeilenumbrüche werden bei *XML*-Anwendungen als *Whitespaces* angesehen und haben für die spätere Darstellung des Textes keine Bedeutung. Sie dienen lediglich der übersichtlicheren Darstellung des *XML*-Textes.²⁴² Der erste Satz des Beispiels könnte genauso gut ohne Zeilenumbrüche und Einrückungen angegeben werden, nur wäre dies deutlich unübersichtlicher:

```
<lb n="1"/>Nachmittags – K. war gerade vor dem Postabschluss sehr
<add hand="Max_Brod" place="inline">134</add> <note>Blaustift</note><lb
n="2"/>beäftigt – drängte sich zwischen zwei Diener <corr>n</corr>, die
Schriftstücke<lb n="4"/><add n="3" place="supralinear" rend="addline">herein</
add>trugen<del type="overstrike">Onkel Karl</del>, K.s Onkel Karl, ein kleiner
Grund-<lb n="5"/>besitzer vom Lande, ins Zimmer.
```

Beispiel 12. Der erste Satz des Kafka-Beispiels ohne *Whitespaces*

Problematisch ist die Kodierung der von Kafka zwischen den Zeilen eingefügten Korrekturen und Ergänzungen. Die zweidimensionale Anordnung des Textes auf dem Papier erfordert bei der Umsetzung in eine lineare *XML*-Struktur einige Überlegungen. Ich habe mich dafür entschieden, die betreffenden Stellen an der Position im Text einzufügen, an die sie aller Wahrscheinlichkeit nach gehören. Im Sinne des Konzeptes der FKA mag diese Entscheidung strittig sein, geht sie doch von der Annahme aus, dass sich alle Einfügungen Kafkas eindeutig einer Textzeile zuordnen lassen, was als interpretierender Eingriff in den Text der Handschrift gewertet werden könnte. Stellen, an denen Kafka die Position des eingefügten Textes mit Hilfe einer Einfügelinie bestimmt hat oder an denen ein ergänztes Wort unmittelbar über einem gestrichenen steht, sind diesbezüglich weniger kritisch als Stellen, an denen die genaue Zuordnung erschlossen werden muss. Meine Entscheidung stellt einen Kompromiss zwischen Kodierungsaufwand und Informationsverlust dar. Alle Einfügungen habe ich mit dem Element »add« gekennzeichnet. Das Element »add« habe ich um die *Attribute* »place«, »n« und »rend« ergänzt. Das »place«-Attribut gibt an, wo die Einfügung in Bezug auf die Textzeile steht. Der Wert »infralinear« steht für eine Einfügung innerhalb der jeweiligen Zeile; »supralinear« für eine Einfügung, die oberhalb der Zeile angegeben ist. Im Fall einer supralinearen Einfügung habe ich zusätzlich das Attribut »n« verwendet, um die Nummerierung der Zwischenzeilen, wie sie die differenzierte Umschrift vornimmt, zu übernehmen. Das

242 Vgl. *Whitespaces*, S. 86.

Attribut »rend« mit dem Wert »adline«²⁴³ habe ich schließlich dazu benutzt, um zu kennzeichnen, wenn die Position der Einfügung durch eine Einfügelungsline markiert ist.

Eine Darstellung, die exakt der differenzierten Umschrift entspricht, ist mit der vorgeschlagenen Kodierung unter Zuhilfenahme geeigneter Technologien zur Formatierung, wie XSL, nicht möglich.

gegluht, wie er ~~ein wenig gebücht~~ den eingedrückten Paramahut
 in der Linken, die Rechte schon von weitem ihm
 entgegenstreckte und ~~sie~~ mit rücksichtsloser Eile über
 den Schreibtisch hinreichte, ~~rücksichtslos~~ alles umstoßend, was ihm im Wege war ~~auf dem bei dieser~~
~~Gelegenheit etwas umfiel~~ rechts und links ~~der~~ ~~oder gar~~ Der Onkel ~~war~~ ~~befand~~
 immer in Eile, denn er war von dem ~~unglücklichen~~ Gedanken
 verfolgt bei seinem ~~stummer nur~~ eintägigen Aufenthalt in der
 Hauptstadt müsse er alles erledigen können, was er
 sich vorgenommen hatte und ~~dürfe~~ überdies ~~auch~~ kein gelegentlich
 sich anbietendes Gespräch oder Geschäft oder Vergnügen
 sich entgehen lassen. Dabei mußte ihn ~~aber~~ K., der

Abbildung 29. Darstellung eines Auszuges des TEI-kodierten Kafka-Beispiels mit einem einfachen XSL-Stylesheet

Eine lineare Darstellung der Streichungen und Einfügungen, wie sie Abbildung 29 zeigt, ist dagegen problemlos zu realisieren.²⁴⁴

Um eine der Buchausgabe ähnliche Darstellung zu ermöglichen, müsste der Schwerpunkt der Kodierung geändert werden. Dabei müssten Aspekte der Textstruktur zugunsten von Darstellungsaspekten in den Hintergrund treten. Auch elektronische Editionen befreien den Herausgeber nicht von grundsätzlichen edito-
 rischen Entscheidungen wie der, auf welchen Aspekten der Schwerpunkt einer Edition liegt.

Die Nummerierung der einzelnen Zeilen hat auch im Kontext elektronischer Texte den Vorteil, dass sich die Zeilen adressieren lassen. In Verbindung mit den Angaben der Seitenumbrüche ist es möglich, eine aus Büchern gewohnte Art des Textnachweises zu verwirklichen. So kann mit Hilfe von Technologien wie XPath und XPointer aus anderen elektronischen Texten auf Stellen in der elektronischen Edition verwiesen werden. Damit ließe sich die ursprüngliche Idee von Hypertexten verwirklichen,²⁴⁵ da nicht mehr für jeden Verweis eine Zielmarkierung im entsprechenden Zieltext eingefügt werden muss, wie es bei Hyperlinks, die mit Hilfe von HTML realisiert wurden, notwendig ist.

Um mit Hilfe von HTML einen Hyperlink zu realisieren, muss im Ausgangsdokument mit Hilfe des Anker Elements »a« ein Ziel angegeben werden:

In der `ersten Zeile` des Kapitels...

243 Ein Problem bei der Textauszeichnung nach TEI-Vorgaben ist, dass die Richtlinien in einigen Fällen Attributwerte vorgeben, etwa für das Attribut »place« die Werte »infralinear«, »superlinear« und »supralinear«; in anderen Fällen bleibt es dem Anwender überlassen, geeignete Werte zu formulieren. Hierbei muss man sich entscheiden, ob man aus Gründen der Einheitlichkeit englische Termini verwendet oder auf deutsche zurückgreift. Ich habe mich dafür entschieden, die Textauszeichnung weitgehend in Englisch zu formulieren, konnte aber kein adäquates Wort für »Einfügelungsline« finden und habe deshalb den Begriff »adline« erfunden.

244 Die Darstellung des TEI-kodierten Textes wird durch ein Stylesheet gesteuert (vgl. Technologien zur Formatierung, S. 92). Die abgebildete Darstellung ist nur ein einfaches Beispiel.

245 Vgl. 1.4.3 Hypertext und Hypermedia, S. 22.

Im Zieldokument muss an der entsprechenden Stelle das Sprungziel eingefügt werden. Dazu dient in *HTML* ebenfalls das Anker-Element:

```
<a name="ziel_1">Nachmittags – K. war gerade vor dem Postabschluss sehr...
```

Gesetzt den Fall, dass viele Autoren mit dieser Technik auf Stellen in einem *HTML*-Dokument verweisen wollen, müsste allen die Möglichkeit eingeräumt werden, an den Stellen, auf die sie verweisen wollen, Sprungziele einzufügen. Dass hieße zum einen, dass alle Autoren in den *HTML*-Quelltext der Edition eingreifen müssen, um die von ihnen benötigten Sprungziele einzufügen; zum anderen würde ein solcher Text unter Umständen mit Sprungzielen gespickt sein. Auf den *TEI*-kodierten Beispieltext ließe sich mit Hilfe von *Xpath* verweisen, ohne dass in den *TEI*-kodierten Text eingegriffen werden müsste:

```
In der <link ziel="pb[ed='Der Onkel Leni' @n='1']/lb[@n='1']">
ersten Zeile</link> des Kapitels...246
```

Da die Adressierung des Ziels komplexer ist, weil nicht ein eingepflegtes Sprungziel als Referenz dient, sondern die Dokumentstruktur, braucht nicht in das Zieldokument eingegriffen zu werden. Ein weiterer Vorteil der hier vorgeschlagenen Referenzierung ist, dass das Ziel bei entsprechender Kenntnis der *TEI*-Kodierung auch ohne eine Software, die die Angaben benutzerfreundlich umsetzt, erkennbar ist: Das Ziel ist auf der Seite (pb) der Edition (ed) »Der Onkel Leni« mit der Nummer (n) 1 in der Zeile (lb) 1.

Für Korrekturen innerhalb der Zeile habe ich im obigen Beispiel das Element »corr« (correction, Zeilen 15, 23, 43, 73) verwendet. Das »corr«-Element schließt dabei die ergänzten bzw. korrigierten Zeichen ein. Im Falle einer Korrektur ist die ursprüngliche Zeichenfolge im »sic«-Element angegeben (Zeilen 23 und 72). Optional können diese Elemente auch um das Attribut »hand« ergänzt werden, in dem angegeben werden kann, von welcher Hand eine Korrektur vorgenommen wurde. Für das obige Beispiel habe ich die Voraussetzung angenommen, dass alle Veränderungen von Kafka stammen, sofern sie nicht um das »hand«-Attribut ergänzt werden.

Für Ergänzungen, die über eine einfache Korrektur hinausgehen, habe ich das Element »add« (Addition) verwendet (s. o.). Gleich das erste Vorkommen (Zeile 13) weist eine Besonderheit auf: Die Seitenzählung am rechten Seitenrand wurde von Max Brod eingefügt. Das »add« Element ist deshalb um das Attribut »hand="Max Brod"« ergänzt. Dem Attribut »place« habe ich hier den Wert »inline« gegeben. Alternativ wäre zu überlegen, ob man für die eingefügten Seitennummern nicht generell eine eigene Positionsbezeichnung, etwa »rechte obere Blattecke« einführt. Das »add«-Element kann auch komplexere Informationen enthalten. Je nachdem, wie ausführlich die Eigenheiten der Handschrift kodiert werden sollen, könnte die erste Ergänzung Kafkas (Zeile 18) neben der Beschreibung der Position des eingefügten Textes über der Zeile auch noch durch die Angaben der Hand, die die Ergänzung eingefügt hat, und mit Einweisungslinie oder eigener Zeilennummer ergänzt werden und sähe dann etwa so aus:

```
<add place="supralinear"
hand="Franz_Kafka"
rend="Einweisungslinie"
n="3">herein</add>trugen
```

Das Attribut »rend«, rendering, macht Angaben, wie der Text in der Vorlage »wiedergegeben« wird bzw. dargestellt ist.

Getilgter Text wird durch das Element »del«, deletion, ausgezeichnet. Auch hier können verschiedene *Attribute* ergänzt werden. Verpflichtend ist gemäß dem *TEI*-Schema das Attribut »type«, in dem näher bestimmt werden muss, um welche Art von Tilgung es sich handelt. Auch hier empfehlen sich wieder klar definierte

²⁴⁶ Es handelt sich hier lediglich um Beispiele, die die prinzipiellen Möglichkeiten verdeutlichen sollen. Ohne weiteren Kontext »funktionieren« diese Beispiele nicht.

Begriffe, ich verwende für durchgestrichenen Text »overstrike« (z. B. Zeile 19). Ergänzend kann die Angabe der Streichung von fremder Hand durch das Attribut »hand« hinzukommen (Zeile 78).

Zuletzt habe ich das Element »note« für Anmerkungen verwendet, die in der Buchausgabe in den Fußnoten stehen (Zeilen 14, 77, 79). Alternativ könnte hierzu auch das »rend« Attribut verwendet werden. Welche Art der Auszeichnung angebracht ist, lässt sich nicht pauschal sagen. Sollen *Attribute* später zur automatischen Verarbeitung, etwa zur Steuerung der Darstellung, herangezogen werden, wäre es von Vorteil, die zulässigen Attributwerte, also die hier verwendeten Termini, klar zu definieren und auf eine überschaubare Anzahl zu begrenzen. Greift man auf diese Methode der Darstellungssteuerung zurück, muss man ggf. später für jeden möglichen Attributwert eine Bedingung definieren, wie im Falle der jeweiligen Attributwerte zu verfahren ist. Um das »rend«-Attribut zu Formatierungszwecken heranziehen zu können, wäre es vorteilhaft, wenn der Attributwert beispielsweise jeweils »violett« hieße. Ein Programm, das den Text formatiert darstellt, könnte in diesem Fall die folgende Anweisung bekommen: Wenn das Attribut »rend« den Wert »violett« hat, dann stelle den Text, den das »add« Element einschließt, violett dar! Diese eindeutige Anweisung ist schwieriger zu formulieren, wenn statt des festgelegten »violett« heterogene Bezeichnungen vorkommen, wie »violette Tinte«, »mit violetter Tinte« etc. Sollen komplexere Angaben gemacht werden, wie dies bei Anmerkungen üblich ist, empfiehlt sich die Verwendung eines eigenen Elements, in diesem Fall des »note«-Elements.

An diesem Beispiel ließen sich noch viele Aspekte der Textauszeichnung nach den *TEI*-Richtlinien diskutieren: Alternative Kodierungsmodelle sowie mögliche Anknüpfungspunkte zur Darstellungssteuerung, Verlinkung oder zur Realisierung von *Retrieval*-Funktionen. Dazu wäre aber ein vertieftes Wissen in den einzelnen Bereichen notwendig, das weit in den Bereich der Informatik hineinführen würde. Aus diesem Grund möchte ich es bei den dargestellten Aspekten bewenden lassen und zum nächsten Beispiel übergehen.

3.3.5 Beispiel aus Georg Büchner: DANTONS TOD

Das editorische Konzept der *MARBURGER BÜCHNER-AUSGABE* (MBA) kann dem der *FRANKFURTER KAFKA AUSGABE* beinahe als entgegengesetzt bezeichnet werden. Während die FKA versucht, keinen »endgültigen« Text zu präsentieren, präsentiert die MBA nicht weniger als fünf Varianten des Textes, zuzüglich zweier Bände zur Dokumentation der Quellen Büchners und der Entstehungsgeschichte des Textes.

Die im Auftrag der Akademie der Wissenschaften und der Literatur in Mainz herausgegebene Marburger Büchner-Ausgabe²⁴⁷ beinhaltet unter anderem die Hauptwerke Büchners. Die Texte werden, sofern möglich, im Faksimile der Handschrift und ihrer differenzierten Umschrift sowie in Form eines genetischen bzw. synoptischen, eines emendierten und eines quellenbezogenen Textes dargestellt. Dazu kommen umfangreiche Apparate mit Angaben über historische Quellen und weiteren Erläuterungen zu den Texten. Den vielfältigen Ansprüchen an den Text und dem hohen Anspruch an die Edition entsprechend fällt die typografische Darstellung der Texte sehr komplex aus, was nicht ohne Kritik blieb:

Historisch-kritische Ausgaben sind Werke der Wissenschaft. Jeder Neuling, der sie zum ersten Mal aufschlägt und unmittelbare Belehrung erhofft, muß erst einmal das Lesen lernen. Anders als im Konversationslexikon bleibt auch dem Geübten eine Ordnung lange Zeit undurchschaubar; wie in einer großen Bürokratie wird er mit einer Frage von einem Band zum anderen geschickt. Im Zeitalter der Computeredition hat sich diese Neue Unübersichtlichkeit eher noch verstärkt: Zu verlockend sind die Möglichkeiten vielfacher Schriftauszeichnung, einer veränderten Buchstabengröße, der Unterstreichung, eingefügter Sonderzeichen. Die Marburger Ausgabe möchte hinter diesen Standard nicht zurückfallen und belehrt den Anfänger, daß Danton keineswegs einen einfachen Tod gestorben ist. Vier Fassungen des vermeintlich einen Dramas kommen in den Bänden zum Abdruck, aus dem Sterben wird ein Massaker.²⁴⁸

²⁴⁷ Büchner, Sämtliche Werke und Schriften.

²⁴⁸ Wirtz, Thomas: Stirb viermal, jetzt erst recht. Die Kompanie »Marburg« der Büchner-Forschung zog ins Feld und kehrt mit vollen Schulranzen zurück. Frankfurter Allgemeine Zeitung Nr. 289, 12.12.2000.

Die Marburger Büchner Ausgabe wurde als reine Buchausgabe konzipiert und bedient sich der umfangreichen Möglichkeiten einer Buchedition: Extensiv werden Typografie, Apparate, Fußnoten und Lemmata genutzt, um die auf mehrere Bücher aufgeteilten Informationen des Bandes zu verknüpfen. Bei der Arbeit an ihr kamen zweifellos elektronische Mittel und Medien zum Einsatz, da ein solches Editionsprojekt in der heutigen Zeit ohne den umfangreichen Einsatz von Computertechnik nicht mehr denkbar ist.

Ich greife als Beispiel eine Manuskriptseite aus *DANTONS TOD* heraus. Die Auswahl der Textstelle erfolgte in dem Bestreben, ein repräsentatives Beispiel zu finden, das möglichst viele Aspekte der Edition enthält. Ich möchte im Folgenden zuerst die einzelnen Textversionen der Marburger Ausgabe vorstellen und anschließend Möglichkeiten einer *TEI*-konformen elektronischen Umsetzung der Informationen diskutieren.

Faksimile der Handschrift

Das Faksimile der Handschrift ist nicht weiter erläuterungsbedürftig. Es handelt sich um die 40. Seite des Manuskripts. Die Befunde dokumentieren die differenzierte Umschrift.

Differenzierte Umschrift

Der differenzierten Umschrift sind folgende Befunde zu entnehmen:

- in Zeile 4 wurde das Wort »kop« von Büchner getilgt, das »p« ist nur ein Buchstabenansatz
- in Zeile 5 wurde der Punkt durch ein Komma überschrieben
- zwischen den Zeilen 8 und 9 wurden zu einem späteren Zeitpunkt die Zeilen R1 bis R5 eingefügt
- das Komma in Zeile R2 wurde durch die Einweisungslinie verdeckt und wieder eingefügt
- in Zeile 12 ist nicht klar, ob der von Büchner getilgte Buchstabe ein »S« oder der Ansatz eines »R« ist
- in den Zeilen 13 und 15 wurde zu einem späteren Zeitpunkt jeweils das »Gut« um ein »e« zu »Gute« ergänzt
- in Zeile 15 wurde später das große »I« gestrichen und durch das kleine »i« in der darüberstehenden Zeile 14 korrigiert
- in Zeile 19 ist die Buchstabenfolge »mme« des Wortes »komme« verschliffen, die einzelnen Striche der Handschrift sind nicht eindeutig den einzelnen Buchstaben zuzuordnen, dennoch kann aufgrund der Eigentümlichkeiten der Handschrift Büchners die Lesung als gesichert gelten
- gleiches gilt in Zeile 28 für die Buchstaben »ac« in »Nachbar« und »mer« in »schmerzt«
- in Zeile 22 ist nicht ersichtlich, ob es sich bei dem getilgten Buchstaben um ein »e« oder den Ansatz eines »v« gehandelt hat
- in Zeile 24 wurde zu einem späteren Zeitpunkt das »bleibe« zu einem »bliebe« korrigiert, indem das »ei« gestrichen und durch ein darüber geschriebenes »ie« ersetzt wurde
- hinzu kommen die Anmerkungen zu den Strichen durch fremde Hände.

Genetische Darstellung

Ergänzend zur differenzierten Umschrift enthält die genetische Darstellung im Wesentlichen folgende Informationen:

- das in Zeile 4 getilgte »kop« sollte wohl ursprünglich »kopf« heißen, das Ausschreiben des Wortes wurde von Büchner abgebrochen und sofort durch das korrekt am Anfang groß geschriebene »Kopf« korrigiert
- in Zeile 12 hat der Herausgeber die Abkürzungen »Ad.« und »u.« als »Adelaide« und »und« ergänzt.

Danton. Jacobins Republik ist ein wahres
 großes Land, wovon wir die Güter
 nicht kaufen antike sind. Sie ist ein
 Königreich, und das ist das Land
 nicht, ist, geht das Land nicht an,
 da alle ist ein Anzeichen, wie
 jeder ein Vorbild sein sollte hat,
 so für das Land die die gut ist.

Lacornis. Jacobins Republik ist ein wahres
 großes Land, wovon wir die Güter
 nicht kaufen antike sind. Sie ist ein
 Königreich, und das ist das Land
 nicht, ist, geht das Land nicht an,
 da alle ist ein Anzeichen, wie
 jeder ein Vorbild sein sollte hat,
 so für das Land die die gut ist.

Danton. Gutes Land, ist für alle Kinder!
Lacornis. Gutes Land, ist für alle Kinder!
 sprechen!

Danton. Sie können mich, sie können
 um die Hand zu sein.

Lacornis. Für Danton, ist keine von
 den Jacobinern.

Danton. Nicht weiter?

Lacornis. Die die man verkaufen eine
 Pantheontion, für man ist alle
 Jahr nicht genug, und sie in die
 Augen zu weichen, jeder muss ein
 Gesetz, und sollte es zu einem
 Kaufman sagen: Danton ist ein Gesetz nicht!

x Lacornis Jacobins
 unempfangige Republik
 jede Kind in einem
 Kindel 2. J. in einem
 neuen Körper.

Abbildung 30. Faksimile der 40. Manuskriptseite von Dantons Tod aus Büchner, Georg: Sämtliche Werke und Schriften. Historisch-kritische Ausgabe mit Quellendokumentation und Kommentar. Bd. 3.1, S. 84

		<i>Differenzierte Umschrift</i>	
			<i>p. 40</i>
		<u>Danton.</u> Fräulein Rosalie ist ein restaurierter Torso, woran nur die Hüften und Füße antik sind. Sie ist eine Magnetnadel, was der Pol [koṗ] Kopf abstößt, zieht der Pol Fuß an[.→], die Mitte ist ein Aequator, wo jeder eine Sublimattaufe nöthig hat, der zum Erstenmal die Linie passirt.	1 2 3 4 5 6 7 8
R1	[× <u>Lacroix.</u> Zwei	[×] Rosalie. Schämen Sie sich, unsere Ohren	9
R2	<u>barmherzige Schwestern</u> [,][,]	roth zu machen!	10
R3	<u>jede dient in einem</u>	<u>Adelaide.</u> Sie sollten mehr Lebensart	11
R4	<u>Spital d. h. in ihrem</u>	haben. ([§]ṙ[→]Ad. u. Rosalie, ab)	12
R5	<u>eigenen Körper.]</u>	<u>Danton.</u> Gut[e] Nacht, ihr hübschen Kinder!	13
		[i]	14
		<u>Lacroix.</u> Gut[e] Nacht, [l]hr Quecksilbergruben!	15 16
		<u>Danton.</u> Sie dauern mich, sie kommen um ihr Nachtessen.	17 18
		<u>Lacroix.</u> Höre Danton, ich komme von den Jacobinern.	19 20
		<u>Danton.</u> Nichts weiter?	21
		<u>Lacroix.</u> Die Lyoner [[e]v] verlasen eine Proclamation, sie meinten es bl[ie]be ihnen nichts übrig, als sich in die Toga zu wickeln. Jeder machte ein Gesicht, als wollte er zu seinem Nachbar sagen: Paetus es schmerzt nicht!	22 23 24 25 26 27 28
	<u>Fremde Hände:</u>	<u>Erläuterungen:</u>	
	li vor Z. 1: waagr. Strich (Blei)	Z. R2: Schwestern[,][,] (früheres Komma durch Einweisungslinie verdeckt)	
	li Z. 6-8: senkr. Strich (Blei)		
	li Z. 15: senkr. Strich (Blei)		

p. 40	1	Danton. Fräulein Rosalie ist ein restau-
	2	rirter Torso, woran nur die Hüften
	3	und Füße antik sind. Sie ist eine
	4	Magnetnadel, was der Pol ^{[kop{t}]f} Kopf
	5	abstößt, zieht der Pol Fuß an[.→],
	6	die Mitte ist ein Aequator, wo
	7	jeder eine Sublimattaufe nöthig hat,
	8	der zum Erstenmal die Linie passirt.
R1		Lacroix. Zwei
R2		barmherzige Schwestern[,] ,
R3		jede dient in einem
R4		Spital d. h. in ihrem
R5		eigenen Körper.
	9	Rosalie. Schämen Sie sich, unsere Ohren
	10	roth zu machen!
	11	Adelaide. Sie sollten mehr Lebensart
	12	haben. (S Ř → Ad.(elaide) u.(nd) Rosalie, ab)
	13	Danton. Gut Nacht, ihr hübschen Kinder! e
	15	Lacroix. Gut Nacht, []hr Quecksilber- 15,14 e i
	16	gruben!
	17	Danton. Sie dauern mich, sie kommen
	18	um ihr Nachtessen.
	19	Lacroix. Höre Danton, ich komme von
	20	den Jacobinern.
	21	Danton. Nichts weiter?
	22	Lacroix. Die Lyoner [e v]verlasen eine
	24	Proclamation, sie meinten es bl[ei]be 23 ie
	25	ihnen nichts übrig, als sich in die
	26	Toga zu wickeln. Jeder machte ein
	27	Gesicht, als wollte er zu seinem
	28	Nachbar sagen: Paetus es schmerzt nicht!

Abbildung 32. Genetische Darstellung der 40. Manuskriptseite von Dantons Tod aus Büchner, Georg: Sämtliche Werke und Schriften. Historisch-kritische Ausgabe mit Quellendokumentation und Kommentar. Bd. 3.1, S. 381 f.

Emendierter Text

Die Marburger Ausgabe verzeichnet im Fußnotenapparat des emendierten Textes Varianten in den Drucken bis 1850: Die Erstveröffentlichung 1835 im Phönix²⁴⁹ (j), die ebenfalls 1835 erschienenen Buchausgabe²⁵⁰ (e) mit Korrekturen Büchners in einzelnen Widmungsexemplaren (e^B, e^S) und die von Ludwig Büchner 1850 herausgegebenen »Nachgelassenen Schriften«²⁵¹ (N).

Die jeder Äußerung einer Person vorangestellte Nummerierung verweist auf die Erläuterungen im vierten Band der Ausgabe.

139 D a n t o n. Fräulein Rosalie ist ein restaurirter Torso, woran nur die Hüften und Füße antik sind. Sie ist eine Magnetnadel, was der Pol Kopf abstößt, zieht der Pol Fuß an, die Mitte ist ein Aequator, wo jeder eine Sublimattaufe nöthig hat, der zum Erstenmal die Linie passirt. 20

140 L a c r o i x. Zwei barmherzige Schwestern, jede dient in einem Spital d. h. in ihrem eignen Körper.

141 R o s a l i e. Schämen Sie sich, unsere Ohren roth zu machen!

142 A d e l a i d e. Sie sollten mehr Lebensart haben. (Adelaide und Rosalie, ab) 25

143 D a n t o n. Gute Nacht, ihr hübschen Kinder!

144 L a c r o i x. Gute Nacht, ihr Quecksilbergruben!

145 D a n t o n. Sie dauern mich, sie kommen um ihr Nachtessen.

146 L a c r o i x. Höre Danton, ich komme von den Jacobinern.

18 D a n t o n.] D a n t o n. O laß das; eN (Bemänglung e^B)
19 Magnetnadel,] Magnetnadel; eN 19 Pol Kopf] Pol-Kopf eN 20 Pol Fuß] Pol-Fuß eN
20f an, die Mitte ist bis die Linie passirt] an. eN (teilw. Restitution e^S, Bemänglung e^B)
22 Schwestern,] Schwestern; eN 22 Spital] Spital, eN 25 haben. (Adelaide] haben. J
(A d e l a i d e eN 25 Rosalie, ab] R o s a l i e ab. eN
27 Quecksilbergruben!] Silbergruben. e (teilw. Restitution e^S, e^B)
Metallgruben. N
29 Höre] Höre, eN 29 Jacobinern] Jakobinern N

147 D a n t o n. Nichts weiter?

148 L a c r o i x. Die Lyoner verlasen eine Proclamation, sie meinten es bliebe ihnen nichts übrig, als sich in die Toga zu wickeln. Jeder machte ein Gesicht, als wollte er zu seinem Nachbar sagen: Paetus es schmerzt nicht!

2 Proclamation,] Proclamation; eN 2 meinten] meinten, eN 4 Paetus] Paetus, eN
4 nicht]] nicht! – eN

Abbildung 33. Emendierter Text der 40. Manuskriptseite von Dantons Tod aus Büchner, Georg: Sämtliche Werke und Schriften. Historisch-kritische Ausgabe mit Quellendokumentation und Kommentar. Bd. 3.2, S. 21 f

Quellenbezogener Text

Der quellenbezogene Text schließlich versucht, den Text auf die von Büchner benutzten Quellen zurückzuführen. Auch hier ist jeder Äußerung eine Nummerierung vorangestellt, die auf die Erläuterungen verweist. Unterstreichungen verweisen auf nachweislich von Büchner benutzte Quellen. Fett gedruckter Text zeigt dabei an, dass Büchner hier wörtlich aus einer Quelle zitiert hat, normal gesetzter Text markiert eine sinngemäße

249 Phönix. Frühlings-Zeitung für Deutschland. Frankfurt am Main: J. D. Sauerländer. Jg. 1835, Nr. 73-77, 79-83; 26.-28., 30. u. 31. März, 2.-4., 6. u. 7. April.

250 Dantons Tod. Dramatische Bilder aus Frankreichs Schreckensherrschaft von Georg Büchner. Frankfurt am Main: J. D. Sauerländer, 1835.

251 Nachgelassene Schriften von Georg Büchner. Hg. von Ludwig Büchner. Frankfurt am Main: J. D. Sauerländer, 1850.

Übernahme von Text aus einer Quelle. Die Abkürzungen in der Marginalspalte verweisen dagegen auf die jeweilige Quelle. Ungesicherte Quellen sind in Klammern angegeben.

Auf die Erläuterungen möchte ich nicht im Einzelnen eingehen. Quellen sind im Beispiel ab Zeile 22 nachgewiesen:

139	D a n t o n .	Fräulein Rosalie ist ein restaurirter Torso, woran nur die Hüften und Füße antik sind. Sie ist eine Magnetnadel, was der Pol Kopf abstößt, zieht der Pol Fuß an, die Mitte ist ein Aequator, wo jeder eine Sublimattaufe nöthig hat, der zum Erstenmal die Linie passirt.	
140	L a c r o i x .	Zwei barmherzige Schwestern, jede dient in einem Spital d. h. in ihrem eignen Körper.	
141	15 R o s a l i e .	Schämen Sie sich, unsere Ohren roth zu machen!	
142	A d e l a i d e .	Sie sollten mehr Lebensart haben. (Adelaide und Rosalie, ab)	
143	D a n t o n .	Gute Nacht, ihr hübschen Kinder!	
144	L a c r o i x .	Gute Nacht, ihr Quecksilbergruben!	
145	20 D a n t o n .	Sie dauern mich, sie kommen um ihr Nachtesen.	
146	L a c r o i x .	Höre Danton, ich komme von den Jacobinern.	(z. Sit Me III 125)
147	D a n t o n .	Nichts weiter?	
148	L a c r o i x .	<u>Die Lyoner</u> verlasen eine Proclamation, sie meinten es bl[<u>a</u>]iebe ihnen nichts übrig, als <u>sich in die Toga zu wickeln</u> . Jeder machte ein Gesicht, als wollte er zu seinem Nachbar sagen: <u>Paetus es schmerzt nicht!</u>	UZ X 90 f., T 196 (No 5 I 191) UZ XII 156

Abbildung 34. Quellenbezogene Textdarstellung der 40. Manuskriptseite von Dantons Tod aus Büchner, Georg: Sämtliche Werke und Schriften. Historisch-kritische Ausgabe mit Quellendokumentation und Kommentar. Bd. 3.2, S. 101

- in Zeile 22 gibt es möglicherweise einen Bezug zu Mercier: *Le nouveau Paris*,²⁵² Band 3, S. 125,
- in Zeile 23 ist Text dem Werk Charles Nodier: *Souvenirs, épisodes et portraits pour servir à l'histoire de la Révolution et de l'Empire*,²⁵³ Band 1, S. 191 entnommen,
- in den Zeilen 23 und 25 ist der Text wörtlich bzw. sinngemäß dem Werk *Unsere Zeit*,²⁵⁴ Band 10, S. 90 und Band 12, S. 156 bzw. in Zeile 22 auch Thiers: *Histoire de la Révolution française*,²⁵⁵ Band 6, S. 196 entnommen.

Textauszeichnung nach TEI

Bei einer *TEI*-gemäßen Auszeichnung des Textes ist es sinnvoll, sich bei der Kodierung an der dramatischen Struktur des Textes zu orientieren, da sie den Nutzern vertraut ist und für die meisten Herangehensweisen die Struktur darstellt, an der man sich in einem dramatischen Text orientiert. Um Ansprüchen, die an eine historisch-kritische Ausgabe gestellt werden, gerecht zu werden, müssen neben der Dramenstruktur auch Details der Handschrift bzw. deren Transkription berücksichtigt werden. Die faksimilierte Handschrift kann in elektronischen Medien lediglich als Bild dargestellt werden.²⁵⁶ Da die diplomatische Umschrift die Befunde der Handschrift prä-

252 Mercier, [Louis Sébastien]: *Le Nouveau Paris*. Paris: Fuchs, Ch. Pougens, Ch. Fr. Cramer, o. J. 6 Bde.

253 Nodier, Charles: *Souvenirs, épisodes et portraits pour servir à l'histoire de la Révolution et de l'Empire*. Paris: Alphonse Levasseur, 1831.

254 [Friedrich, Johann Conrad:] *Unsere Zeit oder geschichtliche Übersicht der merkwürdigsten Ereignisse von 1789–1830*, Bde. I–XIII.

255 Thiers, [Louis] A. [dolphé]: *Histoire de la Révolution française*. 8 Bde. Paris: Lecointe et Durey, 1824–1827.

256 Vgl. Image-Editionen, S. 20.

sentiert, ziehe ich sie als Grundlage der Kodierung heran. Die Textgestalt der anderen Darstellungsweisen leitet sich im Wesentlichen aus der Interpretation der diplomatischen Umschrift ab bzw. ergänzt diese durch weitere Informationen.

Neben den *TEI*-Elementen, die in allen Texten vorkommen können (elements available in all *TEI*-documents), müssen für eine Kodierung, die dem Informationsgehalt der Buchausgabe annähernd gerecht wird, weitere Elemente aus *TEI*-Untergruppen herangezogen werden. Da es sich bei dem Text um ein Drama handelt, werden auch die Elemente zur Dramenstrukturierung (performance textes) benötigt. Hinzuzuziehen sind für die Kodierung der Befunde der Handschrift die Elemente zur Transkription von Manuskripten (representation of primary sources). Zur Kodierung der in der Ausgabe verzeichneten Varianten bis 1850 bieten sich die Elemente zur Kodierung eines kritischen Apparates (critical apparatus) an.

Die für die Textauszeichnung maßgebliche Struktur ist die des Dramas. Neben dem *TEI*-Instrumentarium zur Kodierung dramatischer Schriften soll die Kodierung auch der Handschrift gerecht werden. Auch wenn die einzelnen Aspekte, die Kodierung von Dramen oder von Manuskripten getrennt betrachtet, nicht kompliziert erscheinen, ergeben sich aus der Kombination von beiden Probleme. Diese bestehen hauptsächlich darin, dass, wenn man den Text als Drama betrachtet, andere Strukturen relevant sind, als wenn man den Schwerpunkt auf die Transkription der Handschrift legt. Beispielsweise sind die Struktur bildenden Elemente »div«, wenn man den Text als Drama betrachtet Akte, »div type="akt"« und Szenen »div type="szene"«. Die weitere Gliederung erfolgt in Regieanweisungen »stage« und Sprachäußerungen »sp«, die wiederum aus der Angabe des Sprechers »speaker« und dem gesprochenen Text besteht. Der gesprochene Text ist wiederum in Absätze »p« gegliedert und kann zudem Regieanweisungen »stage« enthalten.

Die Gesamtstruktur aus Akten und Szenen lässt sich an dem von mir ausgewählten Beispiel nicht verdeutlichen, da der Textausschnitt zu klein ist. Ich gehe aber davon aus, dass dem Leser die grundlegende Struktur von Dramen vertraut ist und ich deshalb im folgenden Beispiel Auslassungszeichen in Form eines Kommentar-Tags `<!-- ...-->` verwenden kann, um die Struktur zu verdeutlichen. Die *Tags* habe ich zur Veranschaulichung hervorgehoben.

```
<div type="Akt" n="1">
  <head>I. Act.</head>
  <!-- ... -->
  <div type="Szene" n="5">
    <stage>Ein Zimmer</stage>
    <!-- ... -->
    <stage>Lacroix, Adelaide, Rosalie, treten ein.</stage>
    <!-- ... -->
    <sp><speaker>Danton.</speaker><p>Fräulein Rosalie ist ein restaurirter Torso,
    woran nur die Hüften und Füße antik sind. Sie ist eine Magnetnadel, was der Pol
    Kopf abstößt, zieht der Pol Fuß an, die Mitte ist ein Äquator, wo jeder eine
    Sublimattaufe nöthig hat, der zum Erstenmal die Linie passirt.</p></sp>
    <sp><speaker>Lacroix.</speaker><p>Zwei barmherzige Schwestern, jede dient in
    einem Spital d.h. in ihrem eignen Körper.</p></sp>
    <sp><speaker>Rosalie.</speaker><p>Schämen Sie sich, unsere Ohren roth zu
    machen!</p></sp>
    <sp><speaker>Adelaide.</speaker><p>Sie sollten mehr Lebensart
    haben!<stage>(Adelaide und Rosalie ab.)</stage></p></sp>
    <sp><speaker>Danton.</speaker><p>Gute Nacht, ihr hübschen Kinder!
    </p></sp>
    <sp><speaker>Lacroix.</speaker><p>Gute Nacht, ihr Quecksilbergruben!
    </p></sp>
    <sp><speaker>Danton.</speaker><p>Sie dauern mich, sie kommen um ihr
```

```

Nachtessen.</p></sp>
<sp><speaker>Lacroix.</speaker><p>Höre Danton, ich komme von den
Jacobinern.</p></sp>
<sp><speaker>Danton.</speaker><p>Nichts weiter?</p></sp>
<sp><speaker>Lacroix.</speaker><p>Die Lyoner verlassen eine Proclamation, sie
meinten es bliebe ihnen nichts übrig, als sich in die Toga zu wickeln. Jeder
macht ein Gesicht, als wollte er zu seinem Nachbar sagen: Paetus es schmerzt
nicht!<!-- ... -->
</p></sp>
<!-- ... -->
</div>
<!-- ... -->
</div>

```

Beispiel 13. Dramenorientierte XML/TEI-Kodierung der 40. Manuskriptseite aus Georg Büchner: DANTONS TOD

Im Gegensatz dazu stehen unter dem Aspekt der Auszeichnung von Manuskripten andere Strukturmerkmale im Vordergrund, etwa die Seiten- und Zeilenumbrüche der Handschrift, Einfügungen, Korrekturen etc:

```

<div type="Manuskriptseite" n="40">
  <p>
    <lb n="1" ed="hs"/>
    <hi rend="underlined">Danton</hi>. Fräulein Rosalie ist ein restau-
    <lb n="2" ed="hs"/>rirter Torso, woran nur die Hüften
    <lb n="3" ed="hs"/>und Füße antik sind. Sie ist eine
    <lb n="4" ed="hs"/>Magnetnadel, was der Pol
    <del rend="strikedout">ko
      <unclear reason="Buchstabenansatz">p</unclear>
    </del>
    Kopf
    <lb n="5" ed="hs"/>abstößt, zieht der Pol Fuß an
    <sic>.</sic><corr>,</corr>
    <lb n="6" ed="hs"/>die Mitte ist ein Äquator, wo
    <lb n="7" ed="hs"/>jeder eine Sublimattaufe nöthig hat,
    <lb n="8" ed="hs"/>der zum Erstenmal die Linie passirt.
  </p>
  <p>
    <add place="links">
      <lb n="R1" ed="hs"/>
      <hi rend="underlined">Lacroix</hi>. Zwei
      <lb n="R2" ed="hs"/>barmherzige Schwestern
      <gap reason="früheres Komma durch verdeckt"/>,
      <lb n="R3" ed="hs"/>jede dient in einem
      <lb n="R4" ed="hs"/>Spital d.h. in ihrem
      <lb n="R5" ed="hs"/>eignen Körper.
    </add>
  </p>
  <p>
    <lb n="9" ed="hs"/>
    <hi rend="underlined">Rosalie</hi>. Schämen Sie sich, unsere Ohren
    <lb n="10" ed="hs"/>roth zu machen!
  </p>
  <p>
    <lb n="11" ed="hs"/>
    <hi rend="underlined">Adelaide</hi>. Sie sollten mehr Lebensart
    <lb n="12" ed="hs"/>haben! (
    <del rend="strikedout">
      <unclear rend="S oder Ansatz eines R">S</unclear>
    </del>
  </p>

```

```

    <abbr>Ad.<expan>Adelaide</expan></abbr>
    <abbr>u.<expan>und</expan></abbr>
    Rosalie ab.)
</p>
<p>
  <lb n="13" ed="hs"/>
  <hi rend="underlined">Danton</hi>. Gut<add>e</add> Nacht, ihr hübschen
  Kinder!
</p>
<p>
  <lb n="15" ed="hs"/>
  <hi rend="underlined">Lacroix</hi>. Gut<add>e</add> Nacht,
  <sic>I</sic><corr>i</corr>hr Quecksilber
  <lb n="16" ed="hs"/>gruben!
</p>
<p>
  <lb n="17" ed="hs"/>
  <hi rend="underlined">Danton</hi>. Sie dauern mich, sie kommen
  <lb n="18" ed="hs"/>um ihr Nachtessen.
</p>
<p>
  <lb n="19" ed="hs"/>
  <hi rend="underlined">Lacroix</hi>. Höre Danton, ich ko
  <unclear reason="Buchstabenschwünge" resp="Hg">mme</unclear>
  von
  <lb n="20" ed="hs"/>den Jacobinern.
</p>
<p>
  <lb n="21" ed="hs"/>
  <hi rend="underlined">Danton</hi>. Nichts weiter?
</p>
<p>
  <lb n="22" ed="hs"/>
  <hi rend="underlined">Lacroix</hi>. Die Lyoner
  <del rend="stirkedout">
    <unclear rend="e oder Ansatz eines v">e</unclear>
  </del>
  verlassen eine
  <lb n="24" ed="hs"/>Proclamation, sie meinten es bl<sic>ei</sic>
  <corr>ie</corr>be
  <lb n="25" ed="hs"/>ihnen nichts übrig, als sich in die
  <lb n="26" ed="hs"/>Toga zu wickeln. Jeder macht ein
  <lb n="27" ed="hs"/>Gesicht, als wollte er zu seinem
  <lb n="28" ed="hs"/>N
  <unclear reason="Buchstabenschwünge" resp="Hg">ac</unclear>
  bar sagen: Paetus es sch
  <unclear reason="Buchstabenschwünge" resp="Hg">mer</unclear>zt nicht!
  <!-- ... -->
</p>
</div>

```

Beispiel 14. Am Manuskript orientierte XML/TEI-Kodierung der 40. Manuskriptseite aus Georg Büchner: DANTONS TOD

Die obigen Beispiele geben Anlass zur Diskussion. Die Gliederung in Akte, Szenen oder Seiten der Handschrift lässt sich durch die Klassifizierung der Strukturelemente »div« in »Akt« und »Szene« bzw. »Seite« darstellen. Möchte man beide Strukturen in einem Text realisieren, können sich Probleme bei der hierarchischen Strukturierung von Elementen, die die dramatische Struktur abbilden, und Elementen, die Aspekte der Handschrift markieren,

ergeben. Um zu vermeiden, dass Elemente sich überschneiden,²⁵⁷ sollte eine der Strukturen mit Hilfe von leeren Elementen realisiert werden.²⁵⁸ Die Zeilen- und Seitenumbrüche der Handschrift sind in dieser Hinsicht unproblematisch, da die *TEI*-Richtlinien hierfür bereits die Verwendung der leeren Meilenstein-Elemente »pb« (page break) und »lb« (line break) vorsehen.

Weitere Überlegungen erfordert die Frage, wie die Sprecher kodiert werden sollen. In der Handschrift sind die Sprecher durch Unterstreichungen gekennzeichnet. Die Unterstreichung schließt einen, den Sprecher abschließenden Punkt nicht ein. Für die Textdarstellung mit Hilfe eines *Stylesheets*²⁵⁹ kann es aber sinnvoll sein, den abschließenden Punkt in das »speaker«-Element einzubeziehen, beispielsweise wenn man eine Darstellungsform wählt, bei der der Sprecher auf einer eigenen Zeile steht.

```
<sp><speaker>Danton.</speaker><p>Nichts weiter?</sp>
```

Ist der Zeilenumbruch an das Element »speaker« gebunden, ergibt sich bei der Darstellung kein Problem:

Danton.
Nichts weiter?

Richtet sich die Auszeichnung dagegen nach der Unterstreichung im Manuskript, müsste der Punkt außerhalb des Elements stehen, das den Sprecher beinhaltet. In diesem Fall habe ich das Element »hi«, das eine Hervorhebung kennzeichnet, verwendet:

```
<p>
  <lb n="21" ed="hs"/>
  <hi rend="unterstrichen">Danton</hi>. Nichts weiter?
</p>
```

Würde man die gleichen Formatierungen an das »hi«-Element koppeln, fiel die Darstellung anders aus:

Danton
. Nichts weiter?

Umgekehrt lässt nur die manuskriptbezogene Auszeichnung die korrekte Darstellung des handschriftlichen Befundes zu; zudem sieht eine Darstellung, in der Text und Sprecher auf einer Zeile liegen, bei dieser Form der Kodierung besser aus:

Danton. Nichts weiter?

gegenüber

Danton. Nichts weiter?

Um sowohl der Vorlage gerecht zu werden als auch bei der Darstellungssteuerung flexibel zu sein, ist es in diesem Fall angebracht, die dramen- und die manuskriptorientierte Auszeichnung zu kombinieren.

Ein Problem, das sich aus Berücksichtigung der dramen- und der manuskriptbezogenen Textauszeichnung nach *TEI*-Richtlinien ergibt, sind die unterschiedlichen Empfehlungen zur Verwendung des Absatz-Elements »p«. Bei der Kodierung von Dramen empfehlen die *TEI*-Richtlinien, die einzelnen Absätze eines gesprochenen Textes mit dem Absatz-Element »p« auszuzeichnen; der mit »speaker« ausgezeichnete Sprecher steht außerhalb des Absatz-Elements. Bei der Darstellung von Dramen steht der Sprecher jedoch meist am Beginn des ersten Absatzes. Demnach wäre es folgerichtig, den Sprecher innerhalb des ersten Absatz-Elements einzufügen. Diese

²⁵⁷ Vgl. Struktur, S. 83.

²⁵⁸ Vgl. Elemente, S. 82.

²⁵⁹ Vgl. Technologien zur Formatierung, S. 92.

formal schlüssigere Struktur habe ich im Falle der manuskriptbezogenen Auszeichnungsvariante berücksichtigt, für die dramenbezogene Textauszeichnung lassen die *TEI*-Vorgaben dies nicht zu. Diese Struktur erschwert eine formatierte Darstellung mit Hilfe der zur Verfügung stehenden Technologien, da Absatzmarkierungen bei der Textdarstellung in den meisten Fällen als Blockelemente verwendet werden, die einen eigenen Absatz bilden. Damit müsste der Sprecher auf einer eigenen Zeile stehen.

Die angeführten Darstellungsprobleme lassen sich in den meisten Fällen unter Zuhilfenahme von *XSL* lösen, wobei jedoch je nach Kodierungsmodell ein höherer Aufwand bei der Erstellung der Transformationsanweisungen in Kauf genommen werden muss.

Das folgende Beispiel stellt den Versuch dar, soviel Informationen wie möglich mittels der *TEI*-konformen Kodierungsmöglichkeiten zu realisieren.

Ausgangspunkt sind die beiden oben vorgestellten Möglichkeiten der quellenbezogenen und der dramenbezogenen Textauszeichnung (Beispiel 13 und Beispiel 14). Dazu war es notwendig, die Zeilenzählung, die sich in der manuskriptbezogenen Textdarstellung der Gliederung in Absätze unterordnet, aus der hierarchischen Struktur herauszunehmen und mit Hilfe leerer Meilenstein-Elemente zu realisieren, so dass sie weitgehend unabhängig von anderen Struktur bildenden Elementen verwendet werden können.

Einen Schwerpunkt der Marburger Büchner-Ausgabe stellt die Dokumentation der historischen Quellen, auf die Büchner beim Schreiben des Dramas zurückgegriffen hat, dar. Das entsprechende Material füllt den kompletten dritten Teilband der Ausgabe. Dem entsprechend beinhaltet der zweite Teilband die quellenbezogene Version des Textes (s. Abbildung 34). Die *TEI*-Richtlinien enthalten kein spezielles Konzept zur Kodierung von Quellenbezügen. Die im quellenbezogenen Text in der Marginalspalte ausgewiesenen Quellenverweise lassen sich mit Hilfe der für Anmerkungen vorgesehenen »note«-Elemente realisieren. Für die Nachweise einzelner Wörter oder Passagen, die im gedruckten Text fett bzw. durch verschiedene Unterstreichungen dargestellt sind, gibt es keinen adäquaten Kodierungsvorschlag. Hier sind verschiedene Möglichkeiten denkbar.

Eine Möglichkeit bestünde darin, die typografische Textauszeichnung der Buchausgabe zu übernehmen. Das Element »hi«, das für hervorgehobene Textstellen vorgesehen ist, lässt sich mit Hilfe des »rend«-Attributs ausreichend spezifizieren, so dass es nicht zu Verwechslungen mit der quellenbezogenen Textauszeichnung kommen kann. Allerdings würde dieses Vorgehen bedeuten, in der elektronischen Edition eine Form der doppelten Kodierung zuzulassen: Das Buch kodiert die Information typografisch, der elektronische Text kodiert wiederum die typografische Kodierung. Diese Möglichkeit ist vor dem Anspruch, ein eigenständiges Konzept einer elektronischen Edition zu verfolgen, zu verwerfen.

Da es zum Konzept der *TEI* gehört, dass Anwender die vorgegebenen Strukturen erweitern können, wäre die Möglichkeit der Wahl, eine entsprechende Erweiterung zu entwickeln. Neben dem passenden Konzept müsste dazu das *TEI*-Schema angepasst und die Erweiterung dokumentiert werden. Dies ist prinzipiell möglich und vorgesehen.

Da die historischen Quellen ihrerseits als eigenständiger Textteil kodiert werden müssen, bestünde eine weitere Möglichkeit darin, den Dramentext von dort aus mit Hilfe einer geeigneten *XML*-Technologie, etwa *XPath* oder *XPointer*,²⁶⁰ auszuzeichnen.

Im Folgenden ein Vorschlag für eine elektronische Kodierung des Beispieltextes:

260 Vgl. Technologien zur Adressierung, S. 93.

```

1 <TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
  </teiHeader>
  <text>
5   <body>
     <div type="Akt" n="1">
       <head>I. Act.</head>
       <!-- ... -->
       <div type="Szene" n="5">
10        <stage>Ein Zimmer</stage>
          <!-- ... -->
          <stage>Lacroix, Adelaide, Rosalie, treten ein.</stage>
          <!-- ... -->
          <pb n="40" ed="hs"/>
          <lb n="1" ed="hs"/>
          <sp xml:id="sp139">
            <speaker><hi rend="underlined">Danton</hi>.</speaker>
            <p>
15              <app>
                <rdg wit="eN" cause="(Bemänglung eB)">
                  Oh laß das;</rdg>
                </app> Fräulein Rosalie ist ein restau-
                <lb n="2" ed="hs"/> rirter Torso, woran nur die Hüften
                <lb n="3" ed="hs"/> und Füße antik sind. Sie ist eine
25                <lb n="4" ed="hs"/>
                <app>
                  <lem wit="hs">Magnetnadel,</lem>
                  <rdg wit="eN">Magnetnadel;</rdg>
                </app> was der
30                <app>
                  <lem wit="hs">Pol
                    <del rend="strikedout"> ko
                      <unclear reason="Buchstabenansatz">p</unclear>
                    </del> Kopf
35                  </lem>
                  <rdg wit="eN">Pol-Kopf</rdg>
                </app>
                <lb n="5" ed="hs"/> abstößt, zieht der
40                <app>
                  <lem wit="hs">Pol Fuß</lem>
                  <rdg wit="eN">Pol-Fuß</rdg>
                </app>
                <app>
                  <lem wit="hs">
45                    an<choice><sic>.</sic><corr>,</corr></choice>
                    <lb n="6" ed="hs"/> die Mitte ist ein Äquator, wo
                    <lb n="7" ed="hs"/> jeder eine Sublimattaufe nöthig hat,
                    <lb n="8" ed="hs"/> der zum Erstenmal die Linie passirt.
                  </lem>
                  <rdg wit="eN"
50                    cause="(teilw. Rest. eS, Bemänglung eB)"> an.
                  </rdg>
                </app>
                </p>
55              </sp>
            <addSpan place="marginleft" spanTo="sp141"/>
            <lb n="R1" ed="hs"/>
            <sp xml:id="sp140">
              <speaker><hi rend="underlined">Lacroix</hi>. </speaker>

```

```

<p>Zwei
  <lb n="R2" ed="hs"/> barmherzige
  <app>
    <lem wit="hs">Schwestern
      <unclear reason="Einweisungslinie" >_</damage>
      <add>_</add>
    </lem>
    <rdg wit="eN">Schwesterni</rdg>
  </app>
  <lb n="R3" ed="hs"/> jede dient in einem
  <lb n="R4" ed="hs"/>
  <app>
    <lem wit="hs">Spital</lem>
    <rdg wit="eN">Spital_</rdg>
  </app> d.h. in ihrem
  <lb n="R5" ed="hs"/> eignen Körper.
</p>
</sp>
<lb n="9" ed="hs"/>
<sp xml:id="sp141">
  <speaker><hi rend="underlined">Rosalie</hi>. </speaker>
  <p>Schämen Sie sich, unsere Ohren
    <lb n="10" ed="hs"/> roth zu machen!
  </p>
</sp>
<lb n="11" ed="hs"/>
<sp xml:id="sp142">
  <speaker><hi rend="underlined">Adelaide</hi>. </speaker>
  <p>Sie sollten mehr Lebensart
    <lb n="12" ed="hs"/>
    <stage>
      <app>
        <lem wit="hs">
          <del rend="strikedout">
            <unclear reason="S oder Ansatz eines A">S
              </unclear>
          </del>
          <choice>
            <abbr>Ad.</abbr>
            <expan>Adelaide</expan>
          </choice>
          <choice>
            <abbr>u.</abbr>
            <expan>und</expan>
          </choice> Rosalie ab.)
        </lem>
        <rdg wit="eN"><lb/>(
          <hi rend="underlined">Adelaide</hi> und
          <hi rend="underlined">Rosalie</hi>, ab)
        </rdg>
      </app>
    </stage>
  </p>
</sp>
<lb n="13" ed="hs"/>
<note type="Quelle"> z. Sit. Me III 125 </note>
<sp xml:id="sp143">
  <speaker><hi rend="underlined">Danton</hi>. </speaker>
  <p>Gut<add>e</add> Nacht, ihr hübschen Kinder!</p>

```

```

120     </sp>
121     <lb n="15" ed="hs"/>
122     <sp xml:id="sp144">
123       <speaker><hi rend="underlined">Lacroix</hi>. </speaker>
124       <p>Gut<add>e</add> Nacht,
125         <choice>
126           <sic>I</sic>
127           <corr>i</corr>
128         </choice>hr
129         <app>
130           <lem wit="hs">Quecksilber
131             <lb n="16" ed="hs"/>-gruben!
132           </lem>
133           <rdg wit="e" cause="(teilweise Restitution eS, eB)">
134             Silbergruben.
135           </rdg>
136           <rdg wit="N">Metallgruben</rdg>
137         </app>
138       </p>
139     </sp>
140     <lb n="17" ed="hs"/>
141     <sp xml:id="sp145">
142       <speaker><hi rend="underlined">Danton</hi>. </speaker>
143       <p>Sie dauern mich, sie kommen
144         <lb n="18" ed="hs"/> um ihr Nachtessen.</p>
145     </sp>
146     <lb n="19" ed="hs"/>
147     <sp xml:id="sp146">
148       <speaker><hi rend="underlined">Lacroix</hi> .</speaker>
149       <p>
150         <app>
151           <lem wit="hs">Höre</lem>
152           <rdg wit="eN">Höre, </rdg>
153         </app> Danton, ich ko
154         <unclear reason="Buchstabenschwünge" resp="Hg">mme
155         </unclear> von
156         <lb n="20" ed="hs"/> den
157         <app>
158           <lem wit="hs">Jacobinern</lem>
159           <rdg wit="N">Jakobinern</rdg>
160         </app>.
161       </p>
162     </sp>
163     <lb n="21" ed="hs"/>
164     <sp xml:id="sp147">
165       <speaker><hi rend="underlined">Danton</hi> .</speaker>
166       <p>Nichts weiter?</p>
167     </sp>
168     <lb n="22" ed="hs"/>
169     <sp xml:id="sp148">
170       <speaker><hi rend="underlined">Lacroix</hi>. </speaker>
171       <note type="Quelle">UZ X 90 f. T 196 </note>
172       <p>Die Lyoner
173         <del rend="strikedout">
174           <unclear reason="e oder Ansatz eines v">e</unclear>
175         </del> verlassen eine
176         <lb n="24" ed="hs"/>
177         <app>
178           <lem wit="hs">Proclamation, </lem>

```

```

    <rdg wit="eN">Proclamation;</rdg>
  </app> sie <app>
    <lem wit="hs">meinten</lem>
    <rdg wit="eN">meinten.</rdg>
  </app> es bl
  <choice><sic>ei</sic><corr>ie</corr></choice>be
  <lb n="25" ed="hs"/> ihnen nichts übrig, als
  <note type="Quelle">(NoS I 191)</note> sich in die
  <lb n="26" ed="hs"/> Toga zu wickeln. Jeder macht ein
  <lb n="27" ed="hs"/> Gesicht, als wollte er zu seinem
  <lb n="28" ed="hs"/>
  N<unclear reason="Buchstabenschwünge" resp="Hg">ac
  </unclear>bar sagen:
  <note type="Quelle">UZ XII 15b </note>
  <app>
    <lem wit="hs">Paetus</lem>
    <rdg wit="eN">Paetus.</rdg>
  </app> es sch<unclear reason="Buchstabenschwünge" resp="Hg">
  mer</unclear>zt
  <app>
    <lem wit="hs">nicht!<lb/></lem>
    <rdg>nicht! – </rdg>
  </app>
  <pb n="41" ed="hs"/>
  <!-- ... -->
</p>
</sp>
<!-- ... -->
</div>
<!-- ... -->
</div>
</body>
</text>
</TEI>

```

Beispiel 15. XML/TEI-Kodierung der 40. Manuskriptseite von Georg Büchner: DANTONS TOD

Die Zeilenzählung der folgenden Erläuterungen folgen dem Beispiel 15. In den meisten Fällen habe ich jeweils nur das erste Vorkommen eines Elements angegeben.

Wie oben erwähnt, habe ich als Grundlage der Kodierung die dramatische Struktur gewählt. Die Strukturelemente »div« teilen den Text in Akte und Szenen (vgl. Z. 6 und 9). In diese Struktur fügen sich die untergeordneten Elemente für eine Sprachäußerung (»sp«, Z. 16), die wiederum aus den untergeordneten Elementen »speaker« (Z. 17) und den einzelnen Absätzen (»p«, paragraph, Z. 18) des gesprochenen Textes besteht. Regieanweisungen (»stage«) können sowohl vor und zwischen den Sprachäußerungen (Z. 10) als auch innerhalb des gesprochenen Textes (Z. 90) stehen. Überschriften (»head«, Z. 7) stehen gesondert innerhalb der »div«-Elemente.

Der quellenorientierten Kodierung wird das Beispiel gerecht, indem die Seiten- und Zeilenumbrüche der Handschrift als leere »pb« bzw. »lb« Elemente eingepflegt wurden. Auch in diesem Fall habe ich den Elementen jeweils *Attribute* beigefügt, die sowohl die Nummerierung (»n«) als auch die der Nummerierung zugrunde liegende Edition (»ed«), hier die Handschrift, angeben. Diese Angaben dienen auch in diesem Fall vornehmlich dazu, die Konkordanz zur Handschrift herzustellen und eine Möglichkeit zu schaffen, Textstellen elektronisch adressieren zu können.

Auch weitere Befunde der Handschrift sind in der *TEI*-Struktur kodiert worden: Die Unterstreichungen der Sprecher sind in ihrer Ausdehnung der Handschrift entsprechend mittels des »hi«-Elements ausgezeichnet (Z. 17). Den Grund dafür, dass ich hier sowohl eine dem Manuskript gemäße formale wie eine der Dramenstruktur konforme semantische Kodierung vorgenommen habe, wurde bereits oben erläutert.

In der Handschrift getilgte Stellen sind mit Hilfe des Elements »del« (deleted, Z. 32) gekennzeichnet. Das Attribut »rend« erlaubt es, die Art der Tilgung – im vorliegenden Beispiel handelt es sich ausschließlich um Streichungen – genauer zu beschreiben.

Korrekturen habe ich ähnlich wie im Kafka-Beispiel mit den Elementen »sic« und »corr« markiert (s. o.). Die *TEI*-Richtlinien empfehlen, bei der gleichrangigen Angabe von falschem und korrigiertem Text die beiden Elemente »sic« und »corr« noch mit dem Element »choice« zu klammern, um zu verdeutlichen, dass hier zwei Lesarten zur Auswahl stehen. Im Rahmen einer historisch-kritischen Ausgabe wäre dies nicht erforderlich, da hier der Schreibprozess, zu dem auch die beim Schreiben entstandenen Fehler und Korrekturen gehören, dargestellt werden soll und damit eine parallele Darstellung beider Angaben berechtigt erscheint. Sinnvoll kann die Verwendung des »choice«-Elements unter Umständen für die elektronische Suche sein. Eine entsprechende Suchroutine könnte so programmiert werden, dass sie bei einer Volltextsuche beide Versionen berücksichtigt. In Zeile 126 wurde »i« im Wort »ihr« von Büchner zuerst groß geschrieben und dann zu einem kleinen »i« korrigiert, was sich im *TEI*-kodierten Text folgendermaßen darstellt:

```
<choice><sic>I</sic><corr>i</corr></choice>hr
```

Bei einer einfachen Volltextsuche werden *Tags* meistens ignoriert. In diesem Fall würde eine Suchroutine das Wort »Ihr« vorfinden, es sollten aber die Wörter »Ihr« oder »ihr« gefunden werden. Bei der technischen Realisierung einer geeigneten Suchfunktion müsste also eine Lösung gefunden werden, um mit diesem Problem umzugehen. Insofern ist es von Vorteil, Stellen, an denen ein solches Problem auftauchen kann, gesondert auszuzeichnen.

Das Element »unclear« (Z. 33) markiert Stellen, an denen der Befund der Handschrift nicht eindeutig ist, aber mit hoher Sicherheit erschlossen werden kann. Die genaue Ursache der Unklarheit ist im Attribut »reason« genauer erläutert. Im obigen Beispiel habe ich diese Erläuterungen gekürzt, um zumindest ein wenig Übersichtlichkeit zu gewährleisten. »Buchstabenschwünge« heißt etwa, dass der grafische Befund nicht genau einzelnen Buchstaben zugeordnet werden kann, die Lesart aber als gesichert gelten kann. Die *TEI*-Richtlinien sehen für solche Fälle sogar vor, den Grad der Wahrscheinlichkeit für eine bestimmte Lesart mit Hilfe des Attributs »cert« (certainty) anzugeben.

Ein weiteres Problem ergab sich durch die dramenkonforme Verwendung der Absatz-Elemente: Das Element »add« zur Auszeichnung von ergänztem Text, das ich im Beispiel 14 für die eingefügte Stelle »Lacroix. Zwei barmherzige Schwestern [...]« verwendet habe, darf nur innerhalb eines Absatzes verwendet werden. Dies ist bei einer Kodierung, die sich an der Dramenstruktur orientiert, nicht möglich, da der ergänzte Abschnitt unter anderem auch das Sprecher-Element beinhaltet. Alternativ bot sich das Element »addSpan« an, das zur Auszeichnung umfangreicher eingefügter Passagen dient (Beispiel 15, Z. 56). Um Überschneidungen mit anderen Elementen zu vermeiden, ist es als leeres Element ausgelegt, das lediglich den Beginn einer Einfügung kennzeichnet.

Die Problematik, Aspekte der Textanordnung auf dem Papier exakt zu kodieren, habe ich bereits anhand des Kafka-Textes erläutert. Die Einordnung der von Büchner ergänzten Aussage ist in diesem Fall anhand des Befundes eindeutig möglich und darf als zulässig gelten.

Bei der Auszeichnung eingefügter Textbereiche mit Hilfe des »addSpan«-Elements muss das Ende des eingefügten Bereichs gesondert gekennzeichnet werden. Dies geschieht mit Hilfe einer eindeutigen Kennzeichnung durch einen »identifier«, kurz »ID«, die in den Text eingefügt werden muss. Im vorliegenden Beispiel bot es sich an, die Nummerierung der einzelnen Äußerungen, die die emendierte Textdarstellung der Buchausgabe vorgibt, zu verwenden. Die Nummerierung habe ich als »xml:id«-Attribut in das jede Äußerung umschließende »sp«-

Element eingefügt, da sie als Referenz für die Erläuterungen aus dem vierten Teilband der Ausgabe dienen soll (s. o). Im Beispieltext kann sie auch herangezogen werden, um das Ende des eingefügten Textbereiches zu markieren, der mit dem Anfang der nächsten, dem regulären Textfluss folgenden Äußerung zusammenfällt.

Aus der differenzierten Umschrift habe ich zum einen die dramatische Struktur des kodierten Textes übernommen; zum anderen die nachgewiesenen Varianten in den Ausgaben bis 1850. Hier kommen die *TEI*-Elemente zur Kodierung eines kritischen Apparats zur Anwendung: Das Element »app« umschließt Textstellen, zu denen Varianten zu verzeichnen sind (Z. 19). Die einzelnen Varianten werden mit Hilfe des Elements »rdg« (reading) eingefügt; das Attribut »wit« (witness) dient dazu, anzugeben, in welchem Textzeugen die angegebene Variante vorkommt. Leider sehen die *TEI*-Richtlinien kein Attribut vor, um eine Variante weiter zu erläutern. Diesem Zweck am nächsten kommt das Attribut »cause«, das ich in Fällen verwendet habe, in denen neben der Angabe des Textzeugen noch weitere Informationen zu einer Variante angegeben waren.

Die Variante, die als die im Kontext der Edition gültige Variante angesehen wird, wird mit Hilfe des Elements »lem« (Lemma) gekennzeichnet (Z. 27):

```
<app>
  <lem wit="hs">Magnetnadel.</lem>
  <rdg wit="eN">Magnetnadel:</rdg>
</app>
```

Die Angabe des Textzeugen durch das von mir eingeführte Kürzel »hs« für Handschrift könnte entfallen, wenn an anderer Stelle, etwa im Header, angegeben wäre, dass diese als Leittext fungiert.

Die Praxis, Varianten in den Text einzufügen und nicht in einem eigenen Variantenverzeichnis beizufügen erscheint unüblich, hat jedoch im Kontext elektronischer Medien Vorteile. Bei der elektronischen Suche wird beispielsweise gleich die richtige Stelle im Text angezeigt und nicht eine Fundstelle in einem Apparat, von dem aus man erst wieder die betreffende Textstelle finden muss. Zudem ist es weniger aufwendig, die Darstellung verschiedener Textzeugen zu generieren. Es muss lediglich angegeben werden, welcher Textzeuge bei der Darstellung angezeigt und welcher ausgeblendet werden soll. Nicht zuletzt wird gewährleistet, dass die Informationen nicht auseinandergerissen werden, wenn lediglich Teile des Dokuments verwendet oder weitergegeben werden.

Die separate Auszeichnung eines »Leittextes« mittels des »lem«-*Tags* ist für die Suche von Bedeutung. Ähnlich wie bei der Auszeichnung von Korrekturen muss eine Software unterscheiden können, ob das Wort Magnetnadel im Text tatsächlich zweimal vorkommt oder ob es bei der Suche lediglich einmal berücksichtigt werden soll.

Ein weiteres Problem bei der Kodierung von Texten für elektronische Editionen stellen Abkürzungen dar. Im Gegensatz zu einem menschlichen Leser, der Abkürzungen in den meisten Fällen richtig interpretiert, sollten Abkürzungen in Texten für elektronische Editionen nach Möglichkeit aufgelöst werden, da sonst die Suchfunktionen nicht richtig greifen können. Im vorliegenden Beispiel kommen Abkürzungen in der Regieanweisung »(Ad. u. Rosalie, ab)« vor. Das Wort Adelaide wird durch »Ad.« abgekürzt und das Wort »und« durch »u.«.

Um Abkürzungen aufzulösen, sehen die *TEI*-Richtlinien die Elemente »abbr« (abbreviation, Abkürzung) und »expan« (expansion, die Auflösung der Abkürzung) vor. Wie im Falle der Korrekturen sollte auch hier mittels des »choice«-Elements gekennzeichnet werden, dass eine der angegebenen Möglichkeiten bei der Darstellung, vor allem aber bei der Suche berücksichtigt werden soll:

```

<choice>
  <abbr>Ad.</abbr>
  <expand>Adelaide</expand>
</choice>
<choice>
  <abbr>u.</abbr>
  <expand>und</expand>
</choice>

```

Auf die Problematik, den quellenbezogenen Text der Buchausgabe adäquat in einer den *TEI*-Richtlinien gemäßen Form zu kodieren, habe ich bereits hingewiesen (s. o.). Insofern sind die Quellenverweise der Marginalspalte mittels des »note«-Elements (Z. 170) im elektronischen Text realisiert. Eine adäquate Kodierung der Quellenbezüge im Text ist prinzipiell möglich, jedoch nicht mit Hilfe der bestehenden *TEI*-Richtlinien.

3.4 Metadaten

Unter Metadaten versteht man Daten, die einen Datenbestand beschreiben. Oft spricht man sehr allgemein von »Daten über Daten«. In einigen Fällen, besonders auf Seiten der Informatik, wird alles zu Metadaten erklärt, was nicht unmittelbar zu den eigentlichen darzustellenden oder auszuwertenden Daten gehört. Dagegen werden beim Umgang mit Büchern oft nur bibliografische Daten, wie sie zur Katalogisierung erhoben werden, als Metadaten betrachtet.

Unter Metadaten versteht man im Allgemeinen (strukturierte) Daten, die Informationen über andere Daten, Objekte, Entitäten etc. enthalten. Im informationswissenschaftlichen und bibliothekarischen Kontext versteht man hierunter Daten, die der Beschreibung von elektronischen Ressourcen dienen. In letzter Zeit geht der Trend jedoch dahin, den Begriff Metadaten auch für Daten und Kataloge in Datenbanken zu verwenden, die nicht-elektronische Ressourcen beschreiben.²⁶¹

Im Hinblick auf elektronische Editionen kommt Metadaten eine besondere Bedeutung zu. Wie ich dargestellt habe, vermitteln Bücher neben den Informationen, die im Text und im Schriftbild des Textes enthalten sind, auch weitere Informationen, die jeder Nutzer mehr oder weniger explizit beim Umgang mit Büchern berücksichtigt. Dies fängt bei der Verfügbarkeit eines Buches an. Ob es mehrfach in Bibliotheken oder Buchhandlungen vorhanden ist oder nur in wenigen Fachbibliotheken eingesehen werden kann bzw. in einer Fachbuchhandlung beim Verlag bestellt werden muss oder ob es ständig ausgeliehen/ausverkauft ist und neu aufgelegt wird, gibt bereits Auskunft über die Verbreitung und Aktualität eines Buches. In Bibliotheken mit einer systematischen Freihandaufstellung lässt der Standort eines Buches Rückschlüsse auf seinen Inhalt zu, in Buchhandlungen die Einordnung innerhalb der Systematik der Buchhandlung. Bei Bibliotheksexemplaren sieht man nicht selten am Zustand, ob sie intensiv genutzt wurden.²⁶² Solche Informationen, die sich aus dem physischen Kontext von Büchern ergeben, stehen bei elektronischen Medien nicht immer zur Verfügung. Elektronische Editionen, die auf einem Datenträger wie einer CD-ROM oder DVD gespeichert sind, können in einer entsprechenden Verpackung wie Bücher behandelt werden und lassen ggf. noch einige vom Umgang mit Büchern her gewohnte Rückschlüsse zu. Elektronische Editionen, die über ein Computernetz zur Verfügung gestellt werden, haben diesen physischen Kontext nicht mehr. Hier kann man sich zwar behelfen, indem man andere Indizien nutzt, um Rückschlüsse auf Art und Inhalt eines Textes zu ziehen. Die »Internetadresse«, (die Domain, beispielsweise »fu-berlin.de«)²⁶³, unter der ein Text im Internet zu finden ist, erlaubt ähnliche Schlüsse über einen elektronischen Text wie bei

²⁶¹ <http://www.d-nb.de/standardisierung/metadaten/metadaten.htm> (18.05.2008).

²⁶² Vgl. z. B. 2.1.4 Buchproduktion und Distribution, S. 55.

²⁶³ Internetadressen werden meist in Form eines sogenannten Uniform Resource Locator (URL) angegeben (<http://www...>). Ursprünglich sah das Konzept des World Wide Web vor, dass sich URLs und die dort hinterlegten Informationen nicht ändern sollten. Dieses Konzept ließ sich jedoch aus technischen und wirtschaftlichen Gründen nicht halten.

Büchern über die Bibliothek oder Buchhandlung, in der ein Buch verfügbar ist. Wird ein Text allerdings an einem anderen Ort gespeichert, geht diese Information unter Umständen verloren. Aus diesem Grund sollte eine »Ursprungsadresse« unbedingt Bestandteil der Metadaten elektronischer Texte sein.

Suchfunktionen, wie sie in elektronischen Medien benutzt werden, verlangen darüber hinaus eine besondere Sorgfalt bei der Erhebung von Metadaten. Bei der Auswertung von Metadaten mit Hilfe elektronischer Suchmöglichkeiten muss die formale Logik der entsprechenden Programme berücksichtigt werden. Trotz neuerer Ansätze bei der Realisierung elektronischer *Retrieval*-Funktionen beruhen die meisten aktuell genutzten Systeme zur Informationssuche darauf, Übereinstimmungen zum Suchausdruck zu finden. Abstraktionsleistungen, wie etwa das Einbeziehen von Synonymen, können die meisten elektronischen Systeme nicht leisten.

In elektronischen Medien, besonders dem *World Wide Web*, hat sich darüber hinaus ein Umgang mit Texten und Informationen etabliert, der es in vielen Fällen erschwert bis nahezu unmöglich macht, Herkunft und Zuverlässigkeit eines hier gefundenen Textes zu bestimmen. Exakte Quellenangaben findet man selten. Wenn eine Quellenangabe in Form einer »Internetadresse«, angegeben wurde, ist nicht gesichert, dass diese noch existiert oder noch die ursprünglichen Daten enthält.

Metadaten werden im Kontext elektronischer Medien auf verschiedenen Ebenen zu unterschiedlichen Zwecken erhoben und genutzt. Sie lassen sich in der Art ihrer Erhebung, ihrem Speicherort, ihrer Verwendung und ihrer Kodierung unterscheiden.

3.4.1 Erhebung von Metadaten

Metadaten können nach konkreten Vorgaben erhoben, frei formuliert oder automatisch generiert werden.

Standardisierte Metadaten werden beispielsweise bei der formalen Katalogisierung von Büchern nach einem Regelwerk wie den Regeln für die alphabetische Katalogisierung (RAK) bzw. den Regeln für die Formalkatalogisierung (RFK) erhoben. Im Kontext internationaler Forschung finden im deutschsprachigen Raum auch zunehmend die Anglo-American Cataloguing Rules (AACR) Verwendung. Diese Regeln geben unter anderem vor, welche Angaben zur Titelaufnahme relevant sind und in welcher Form sie angegeben werden. Damit regeln sie die Erhebung und die Formulierung entsprechender Metadaten, beispielsweise was als Titel eines Werkes aufgenommen werden soll oder in welcher Form der Name des Verfassers anzugeben ist. Die genannten Standards haben in Bezug auf elektronische Medien den Nachteil, dass sie vorwiegend für die Katalogisierung von Druckwerken in Zettelkatalogen ausgelegt sind. Damit berücksichtigen sie weder die Titelaufnahme elektronischer Editionen noch die Auswertung der erhobenen Daten mit Hilfe elektronischer Medien. Durch die Kodierung der nach bibliografischen Standards erhobenen Daten mit geeigneten Datenformaten (s. u.) stehen die erhobenen Daten zunehmend auch einem elektronischen Zugriff zur Verfügung.²⁶⁴

Standardisierte Metadaten dienen meist der Katalogisierung von Büchern in Bibliothekskatalogen. Auch zur Katalogisierung elektronischer Editionen werden entsprechende Metadaten erhoben und in elektronischen Katalogen veröffentlicht. Die Daten werden meist in einer eigenen Katalogdatenbank gespeichert. Der Zugriff auf die elektronische Edition selbst erfolgt meist konventionell über eine Signatur, die auf einen Datenträger verweist oder angibt, wie auf den entsprechenden Textbestand zugegriffen werden kann. Als Beispiel für diese Art der Verwendung standardisierter Metadaten zeigt Abbildung 35 die Bildschirmdarstellung der Katalogdaten der elektronischen Edition *KAFKAS WERKE IM WWW*.

264 Vgl. 3.4.4 Kodierung von Metadaten, S. 131.

	[Elektronische Ressource]
Titel:	Kafkas Werke im WWW : kritische Kafka-Ausgabe des S. Fischer Verlages ...
Sonst. Personen:	Kafka, Franz *1883-1924*
Erschienen:	[Cambridge] : ProQuest Information and Learning Co., 2004-
Umfang:	Online-Ressource.
Anmerkung:	Parallele CD-ROM-Ausg. u.d.T.: Kafkas Werke
Sachgebiete:	18.10 Deutsche Literatur 17.97 Texte eines einzelnen Autors
Signatur:	HB 5 Sr 8076-ERF
Weitere Signatur:	HA 5 Sr 8205-ERF
Standort:	Im Netz der Staatsbibliothek
Bestand:	Zugriff für registrierte Benutzer der SBB
Anmerkung:	REMOTE ACCESS
Link:	http://erf.sbb.spk-berlin.de/han/381834964

Abbildung 35. Bildschirmarstellung des Katalogeintrags der elektronischen Edition KAFKAS WERKE IM WWW im Online-Katalog der Staatsbibliothek zu Berlin (StaBikat)²⁶⁵

Obwohl es sich um einen elektronischen Katalog handelt, enthalten die Daten »Spuren« der Titelaufnahme nach RAK: Der Titelzusatz ist durch einen Doppelpunkt vom Titel getrennt; eckige Klammern kennzeichnen Angaben, die nicht dem Haupttitelblatt entnommen bzw. selbst recherchiert wurden.

Bei den frei formulierten Metadaten handelt es sich in den meisten Fällen um deskriptive Metadaten (s. u.), die in einer Datei zusammen mit dem elektronischen Text selbst gespeichert werden.

Frei formulierte Metadaten werden nicht nach einem übergeordneten Regelwerk erhoben. Sie dienen zur Beschreibung von Informationen, für die es entweder keine standardisierten Regelwerke gibt oder die sich einer standardisierten Beschreibung entziehen. Da es für elektronische Editionen keine Regelwerke gibt, die genau vorgeben, nach welcher Systematik Metadaten erhoben werden sollen, werden die meisten Metadaten frei formuliert. Im Idealfall wird eine eigene Systematik entwickelt, die im Kontext der Daten veröffentlicht wird. Daneben ist es auf dem Gebiet elektronischer Medien gängige Praxis, Metadaten in Datenformaten zu speichern, die konkrete Informationsstrukturen vorgeben bzw. die es ermöglichen, den Metadaten durch Kodierung semantische Informationen beizufügen. Entsprechenden Kodierungsmodellen wie denen der *Dublin Core Metadata Initiative* oder den *TEI-Richtlinien* werden häufig Empfehlungen beigegeben, wie die jeweiligen Daten erhoben und formuliert werden sollen, so dass diese Kodierungsmodelle ein gewisses Maß an Standardisierung vorgeben.²⁶⁶

Automatisch generierte Metadaten werden an verschiedenen Stellen erhoben. Computer registrieren automatisch, wann eine Datei zum letzten Mal gespeichert wurde. Ebenso verwalten Computer die Zugriffsrechte (wer berechtigt ist, die Daten zu lesen, zu ändern oder zu löschen) auf Dateien und speichern die notwendigen Daten. Server in Netzwerken halten darüber hinaus die Zugriffe auf Dateien, die auf ihnen gespeichert sind, fest. Automatisch generierte Metadaten stehen selten im Blickpunkt geisteswissenschaftlicher Forschung, sondern werden meist dem Bereich der Datenverwaltung zugeordnet. Da sie sich unter Umständen mit jedem Zugriff auf eine Datei ändern, werden sie nicht (wie beispielsweise frei formulierte Metadaten) im elektronischen Dokument selber gespeichert.

Automatisch generierte Metadaten werden zwar fast immer erhoben, jedoch selten zur Verfügung gestellt. Im *WWW* werden sie teilweise benutzt um anzuzeigen, wann ein Dokument zuletzt überarbeitet wurde. Die sogenannten »Besucherkähler« greifen auf die Informationen zurück, die der Server über die Anzahl der erfolgten Zugriffe auf eine Datei festhält. Es wäre wünschenswert, solche Daten auch im Zusammenhang mit elektronischen Editionen zu nutzen. Sie könnten beispielsweise im Rahmen rezeptionsanalytischer Untersuchungen von Interesse sein.

²⁶⁵ <http://stabikat.sbb.spk-berlin.de> (18.05.2008).

²⁶⁶ Vgl. 3.4.4 Kodierung von Metadaten, S. 131.

3.4.2 Speicherung von Metadaten

Metadaten beschreiben in den meisten Fällen die jeweilige »Informationsressource«, also den betreffenden elektronischen Text. Sie sind damit ebenso veränderlich bzw. unveränderlich wie der jeweilige Text und sollten mit dem Text zusammen in einem Dokument bzw. einer Datei gespeichert werden. Dies geschieht meist am Anfang der Datei, so dass diese Daten zuerst ausgewertet werden können. Die meisten Konzepte zur Datenkodierung sehen deshalb am Anfang einer Datei einen sogenannten Prolog oder Kopf vor, auch als »Header« bezeichnet, der die wichtigsten Metadaten enthält. Oft werden hier neben deskriptiven Metadaten auch technische Angaben zur Verarbeitung der Datei gemacht, etwa über die zugrunde gelegte *Auszeichnungssprache* oder die *URL*, unter der das der Kodierung zugrunde liegende *Inhaltsmodell* zu finden ist.

Benutzungsbezogene oder administrative Metadaten können sich dagegen häufig ändern – ggf. schon bei jedem Zugriff auf die Datei. Solche dynamischen Metadaten werden nicht in der Textdatei gespeichert, sondern an anderen Stellen, etwa dem »Logfile«²⁶⁷ des jeweiligen Servers, der die Daten beherbergt. Sie stehen dort für eine weitere Auswertung zur Verfügung. Solche Daten werden im Internet an verschiedenen Stellen erhoben. So werten etwa einige Suchmaschinen Zugriffszahlen aus, um ein so genanntes »ranking« vorzunehmen, also zu bestimmen, wie relevant eine Datei innerhalb einer Liste von Suchergebnissen ist. Man geht dabei davon aus, dass die Inhalte einer Datei, die vergleichsweise häufig aufgerufen wird, für die meisten Nutzer eine höhere Relevanz haben als die einer Datei, die selten aufgerufen wird. Die meisten Suchmaschinen verwenden noch deutlich komplexere Algorithmen als Grundlage des rankings als dem hier dargestellten. Der Zugriff auf Informationsressourcen über solche *Retrieval*technologien ist für literaturwissenschaftliche Fragestellungen nur unter sehr speziellen Aspekten von Interesse. Ich gehe deshalb auf diesen Aspekt nicht weiter ein.

3.4.3 Verwendung von Metadaten

Im Bereich des elektronischen Publizierens werden an Metadaten weitere Anforderungen gestellt, die über die Katalogisierung hinausgehen. Prinzipiell lassen sich Metadaten wie folgt aufteilen:²⁶⁸

Metadaten über Informationsressourcen:

- Deskriptive Metadaten (Objektbeschreibung)
- Administrative Metadaten (Verwaltung, Rechte, Nachweise ...)
- Nutzungsbezogene Metadaten (Zugriffshäufigkeit, Art der Nutzung).

Deskriptive Metadaten

Deskriptive Metadaten umfassen zum einen bibliografische Angaben, wie sie etwa für die Katalogisierung erforderlich sind. Daneben sind im Bereich der elektronischen Datenverarbeitung noch technische Angaben erforderlich. Dies sind zum Beispiel das verwendete Datenformat oder der einer Textkodierung zugrunde gelegte Standard. So ist es unter Umständen notwendig, die vorliegenden Daten näher zu spezifizieren, etwa den einer Textdatei zugrunde gelegten Zeichensatz oder im Falle von Bilddaten das angewandte Kompressionsverfahren, eine Farbreferenz oder ähnliches. Auch genauere Angaben zur Datenherkunft können wichtig sein – etwa zur Wahrung des Copyrights von Texten und Daten.²⁶⁹ Eine nähere Erläuterung der Datengewinnung und der dabei eingesetzten Programme kann von Bedeutung sein, um etwa die Zuverlässigkeit von elektronischen Texten zu beurteilen. Bei retrodigitalisierten Texten lässt sich z. B. durch die Kenntnis des bei der Texterfassung einge-

²⁶⁷ Als Logfile wird eine Datei bezeichnet, in der alle Zugriffe auf einen Server bzw. alle Ereignisse, die auf einem Server stattfinden, protokolliert, also »geloggt« werden.

²⁶⁸ Vgl. Schmidt, Ingrid: Modellierung von Metadaten. In: Texttechnologie. Perspektiven und Anwendungen. Hg. von Henning Lobin und Lothar Lemnitzer. Tübingen: Stauffenberg, 2004. S. 143-164. S. 150 ff.

²⁶⁹ Die Rechtslage ist bei elektronischen Texten oft schwierig. So sind teilweise verschiedene Rechte für die Texte selber, ihre Bearbeitung, die elektronischen Rechte an den Texten und das Datenformat zu berücksichtigen.

setzen Verfahrens²⁷⁰ auf die zu erwartende Textqualität schließen. Der Hinweis auf eingesetzte Software kann auch für die Weiterverarbeitung der Daten relevant sein. Auch die Angabe zugrunde gelegter Standards, z. B. bei der Textauszeichnung oder der Indizierung, ist in vielen Fällen nützlich, teilweise sogar erforderlich. Zuletzt sind unter diesem Punkt auch Daten zur elektronischen Datenauswertung aufzuführen. Von zunehmender Bedeutung sind Daten zur automatischen Auswertung elektronischer Ressourcen. Neuere Technologien wie die des Semantic Web benötigen semantische Daten, die den Inhalt der Ressource näher spezifizieren.

Administrative Metadaten

Administrative Metadaten, also Angaben zur elektronischen Verwaltung, sind unter Umständen wichtig, um die Verbreitung und Verwendung von Texten zu dokumentieren und zu kontrollieren. Dies dient zum einen der Quellensicherung, damit der Nachweis der Datenherkunft gesichert ist; zum anderen ermöglichen solche Daten beispielsweise die Regelung der Zugriffs- und Bearbeitungsrechte. Angaben über eine gebührenpflichtige Verwendung der Texte oder Daten eines automatischen Abrechnungssystems lassen sich ebenfalls unter diesen Punkt fassen, wobei diese Daten auch unter die Kategorie nutzungsbezogene Metadaten fallen können.

Nutzungsbezogene Metadaten

Die Auswertung der Nutzungsart und -frequenz der bereitgestellten Ressource kann unter verschiedenen Aspekten von Bedeutung sein. So werten viele Suchmaschinen die entsprechenden Daten aus, um bei der Ergebnisanzeige gleich die mögliche Relevanz der Suchergebnisse anzuzeigen.

Das Protokollieren der Datennutzung ist für Technologien, die semantische Verbindungen verschiedener Daten bzw. die Relevanz von Daten in verschiedenen Kontexten erkennen sollen, von großer Bedeutung.

Schließlich lassen sich aus nutzungsbezogenen Metadaten Rückschlüsse auf die Verbreitung eines Textes ziehen: Informationen, die bei Büchern etwa der Auflagenhöhe zu entnehmen sind.

Neben den eigentlichen Metadaten können weitere Angaben über die Metadaten selbst erforderlich sein, die als Metadaten über Metadaten bezeichnet werden.²⁷¹

Metadaten über Metadaten:

- Methode der Metadatengewinnung
- Ersteller der Metadaten
- Status der Metadaten (z. B. sind Metadaten statisch/dynamisch)
- Struktur der Metadaten (z. B. *TEI*, *DCMI*, *MABxml*)
- Vokabular der Metadaten
- Bezugsgröße der Metadaten.

Metadaten über Metadaten dienen meist der Auswertung der Metadaten selber. So kann es von Interesse sein, wie die Metadaten gewonnen wurden: Ob die Daten »von Hand« nach bestimmten Richtlinien von einem Menschen erfasst wurden, was meist bei bibliografischen Angaben der Fall ist, oder durch automatische Verfahren, etwa die automatische Erhebung einer Zugriffsstatistik auf einen elektronischen Text. Dazu gehört auch, welche Regeln bei der Erfassung der Metadaten angewendet worden sind und wer die Metadaten erhoben hat.

Da sich das Bereitstellen elektronischer Informationsressourcen vielerorts schrittweise vollzieht, ist es gängige Praxis, Zwischenergebnisse zu veröffentlichen. Besonders bei der schrittweisen Aufarbeitung von Archivbeständen und deren elektronischer Erfassung und Bereitstellung sowie bei umfangreichen Retrodigitalisierungsprojekten ist dies der Fall. Hier ist es wichtig zu erfahren, ob die Erfassung der Metadaten als abgeschlos-

²⁷⁰ Vgl. 1.4.7 Retrodigitalisierung, S. 29.

²⁷¹ Vgl. Schmidt (2004), S. 150 ff.

sen betrachtet werden darf oder ob es sich um einen Zwischenstand handelt, sich also noch Veränderungen ergeben werden.

Nutzungsbedingte Daten müssen ggf. dynamisch gehandhabt werden, da sich diese Daten durch die Nutzung verändern.

3.4.4 Kodierung von Metadaten

Zur elektronischen Speicherung müssen die erhobenen Metadaten in ein geeignetes Datenformat überführt werden. Auch in diesem Fall bietet sich die Strukturierung der Daten mit Hilfe von *Auszeichnungssprachen* an.

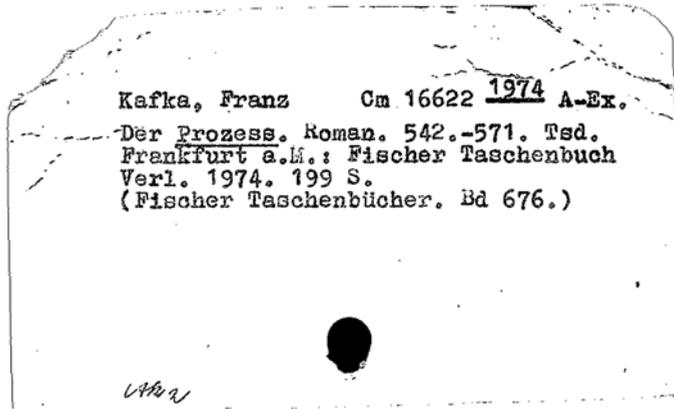


Abbildung 36. Katalogkarte aus einem alphabetischen Zettelkatalog²⁷²

Das Prinzip der Kodierung von Metadaten mit Hilfe einer *Auszeichnungssprache* ist das gleiche wie bei der Auszeichnung von Texten.²⁷³ Die bibliografischen Daten sind auf der Katalogkarte eines Zettelkatalogs, wie sie Abbildung 36 zeigt, durch ihre Anordnung auf der Karte sowie durch Unterstreichungen, Punkte und Klammern gekennzeichnet. Links oben auf der Karte ist der Autor verzeichnet, rechts oben die Signatur bzw. der Standort des Buches. Darunter folgen die weiteren Angaben. Die Bedeutung der Informationen erschließt sich demjenigen, der sich mit Zettelkatalogen auskennt, problemlos. Für ihre Verwendung in elektronischen Medien müssen die Daten jedoch in geeigneter Weise aufbereitet – kodiert – werden. Beispiele dazu führe ich weiter unten an.

Die meisten Kodierungsvorschläge beziehen sich auf die Auszeichnung deskriptiver Metadaten. Hier haben sich verschiedene Standards etabliert.

Maschinelles Austauschformat für Bibliotheken

Das Sortieren bzw. Durchsuchen großer Datenbestände gehört zu den Kernaufgaben der elektronischen Datenverarbeitung. Computer wurden deshalb früh zur Verwaltung von bibliografischen Daten verwendet. Um einen einheitlichen Standard zu schaffen, der den Austausch von Daten zwischen den Katalogen verschiedener Bibliotheken und Datenbanksystemen ermöglicht, entstand 1972 das »Maschinelle Austauschformat für Bibliotheken« (MAB bzw. MAB1),²⁷⁴ das 1995 zu MAB2 weiterentwickelt wurde. Grundlage ist die Katalogisierung nach RAK, wobei die erhobenen Daten in das MAB überführt werden. Das von der amerikanischen Library of

²⁷² Die Katalogkarte ist dem Image Public Access Catalogue (IPAC) der Zentral- und Landesbibliothek Berlin entnommen.

²⁷³ Vgl. 3.1.1 Grundprinzip von Auszeichnungssprachen, S. 71.

²⁷⁴ MAB1 wurde 1972 entwickelt, nachdem die DFG Vorstudien finanziert und die Deutsche Bibliothek sowie die »Arbeitsstelle für Bibliothekstechnik« die Entwicklungsarbeit geleistet hatte.

Congress etablierte Format Machine-Readable Cataloguing (MARC) kann auf internationaler Ebene als Pendant zu MAB betrachtet werden.

Das MAB wird heute in vielen Bereichen der Katalogisierung und des Datenaustauschs zwischen elektronischen Katalogen eingesetzt. Dem Format ist jedoch anzumerken, dass es sich um einen vergleichsweise alten Standard handelt. Die Feldbezeichnungen bestehen aus Nummern, sind also nicht selbsterklärend bzw. unmittelbar verständlich. Dazu kommt die Verwendung verschiedener Sonderzeichen, die als Steuer- oder Trennzeichen eingesetzt werden und heute eher unüblich sind.

Da das MAB ursprünglich entwickelt wurde, um in einem konkret definierten Umfeld Katalogdaten zu verwalten, ist es nicht darauf ausgelegt, Metadaten im Zusammenhang mit elektronischen Texten bzw. Editionen zu speichern. Es handelt sich vielmehr um ein elektronisches Katalogformat. Weil es im deutschsprachigen Raum weit verbreitet ist²⁷⁵ und einen zuverlässigen Standard bietet, wäre es wünschenswert, einer elektronischen Edition die bibliografischen Angaben im MAB beizufügen. Zur Katalogisierung der elektronischen Ressource könnte der Eintrag einfach aus den Metadaten herauskopiert und in einen elektronischen Katalog übernommen werden. In einem nach den *TEI*-Richtlinien kodierten elektronischen Text wäre die geeignete Stelle der *TEI*-Header.²⁷⁶ Um einen bibliografischen Eintrag nach dem ursprünglichen MAB-Standard einzufügen, müssten die Daten in einem CDATA-Abschnitt²⁷⁷ eingefügt werden. Ansonsten könnten durch die von MAB verwendeten Steuerzeichen bei der Verarbeitung mit Programmen, die auf die Verarbeitung von *XML*-Daten ausgelegt sind, Probleme auftreten.

```
<?xml version="1.0"?>
<TEI>
  <teiHeader>
    <!-- ... -->
    <![CDATA[00934nM2.0]200024 h00] 187056-002a19890420-003 20021120131909-
025z187056-030 b|ldcz|z||37-036aDE-037ade-050 |||||g-052 p|||z|||
070 9000-070a292-070b1200-33] Journal of neurology für Testzwecke und
Validation Lokaldaten Heise-335 official journal of the European Neurological
Society-34] Zeitschrift für Neurologie-370aOrgan d. Deutschen Gesellschaft für
Neurologie u. d. Deutschen Gesellschaft für Neurochirurgie- 370aZeitschrift
für Neurologie. 1974-2.-370aInsert for neurologists-376bJNRYA 376bZSNUA 376 J.
Neurol.-376 J. Neurol. (Berlin)-405 207.1974 -4]0 Berlin ; Heidelberg [u.a.]-
4]2 Springer-...]]>
    <!-- ... -->
  </teiHeader>
</TEI>
```

Beispiel 16. Bibliografische Daten im MAB als CDATA-Abschnitt in einen *TEI*-Header eingefügt²⁷⁸

Mittlerweile wurde das MAB₂ auch in *XML* formuliert.²⁷⁹ Dabei wurde aus MAB jedoch kein anwenderfreundlicheres Datenformat, da man an der numerischen Kodierung der Feldbezeichnungen festhielt und die verwendeten Steuerzeichen lediglich in eine *XML*-konforme Form gebracht wurden. Daten im so genannten MAB_{XML}-Format könnten ebenfalls in den Dokumentkopf eines nach den *TEI*-Richtlinien kodierten Dokuments integriert werden. Statt eines CDATA-Abschnittes muss in diesem Fall ein Namespace²⁸⁰ eingerichtet werden:

²⁷⁵ Im englischsprachigen Raum ist das Machine-Readable Cataloguing (MARC) das Pendant zu MAB.

²⁷⁶ Vgl. 3.3.1 *TEI* Richtlinien, Dokumentstruktur, S. 95.

²⁷⁷ Vgl. CDATA-Abschnitte, S. 85.

²⁷⁸ Das Beispiel der MAB-Daten habe ich entnommen: Kett, Jürgen: MABXML 1. Ein XML-Schema für das MAB₂-Format. http://www.ddb.de/standardisierung/pdf/mabxml_1_dok.pdf (18.05.2008). S. 9. Die Beispieldaten habe ich gekürzt.

²⁷⁹ Vgl. Kett, MABXML 1.

²⁸⁰ Vgl. 3.2.5 *XML*-Namensräume, S. 88.

```

<?xml version="1.0"?>
<TEI>
  <teiHeader xmlns:mab="http://www.ddb.de/professionell/mabxml/mabXML
1.xsd">
    <!-- ... -->
    <mab:datensatz typ="h" status="n" mabVersion="M2.0">
      <mab:feld nr="001" ind="n">187056</mab:feld>
      <mab:feld nr="002" ind="a">19890420</mab:feld>
      <mab:feld nr="003" ind="n">20021120131909</mab:feld>
      <mab:feld nr="025" ind="z">187056</mab:feld>
      <mab:feld nr="030" ind="n">b|ldcz|z|||37</mab:feld>
      <mab:feld nr="036" ind="a">DE</mab:feld>
      <mab:feld nr="037" ind="a">de</mab:feld>
      <mab:feld nr="050" ind="n">||||||lg</mab:feld>
      <mab:feld nr="052" ind="n">p|||||z||||||</mab:feld>
      <mab:feld nr="070" ind="n">9000</mab:feld>
      <mab:feld nr="070" ind="a">292</mab:feld>
      <mab:feld nr="070" ind="b">1200</mab:feld>
      <mab:feld nr="331" ind="n">Journal of neurology für Testzwecke
und Validation Lokaldaten Heise</mab:feld>
      <mab:feld nr="335" ind="n">official journal of the European
Neurological Society</mab:feld>
      <mab:feld nr="341" ind="n">Zeitschrift für Neurologie</mab:feld>
      <mab:feld nr="370" ind="a">Organ d. Deutschen Gesellschaft für
Neurologie u. d. Deutschen Gesellschaft für
Neurochirurgie</mab:feld>
      <mab:feld nr="370" ind="a">Zeitschrift für Neurologie.<ns>1974-</ns>
</mab:feld>
      <mab:feld nr="370" ind="a">Insert for neurologists</mab:feld>
      <mab:feld nr="376" ind="b">JNRYA</mab:feld>
      <mab:feld nr="376" ind="b">ZSNUA</mab:feld>
      <mab:feld nr="376" ind="n">J. Neurol.</mab:feld>
      <mab:feld nr="376" ind="n">J. Neurol. (Berlin)</mab:feld>
      <mab:feld nr="405" ind="n">207.1974 -</mab:feld>
      <mab:feld nr="410" ind="n">Berlin ; Heidelberg [u.a.]</mab:feld>
      <mab:feld nr="412" ind="n">Springer</mab:feld>
    </mab:Datensatz>
    <!-- ... -->
  </teiHeader>
</TEI>

```

Beispiel 17. Bibliografische Daten im MABxml Format mit Namespace in einen TEI-Header eingefügt²⁸¹

Der größte Vorteil des MABxml-Formats gegenüber dem herkömmlichen MAB-Format dürfte darin bestehen, dass das *Inhaltsmodell* eines MABxml-Dokuments in Form eines *XML-Schemas* festgelegt wurde. Damit stehen für die Arbeit mit MABxml-Daten viele Programme zum Umgang und zur Prüfung der Daten zur Verfügung.

Die Dublin Core Metadata Initiative

Die Empfehlungen der *Dublin Core Metadata Initiative (DCMI)*²⁸² wurden 1995 zur Kategorisierung von Informationsressourcen im *World Wide Web* entwickelt. Die *DCMI* besteht aus einer Reihe von Arbeitsgruppen, die sich unter dem Dach der *DCMI* zusammen gefunden haben, um einheitliche Standards für Metadaten zu schaffen. Im Kontext dieser Arbeit sind die Bestrebungen der *DCMI* interessant, die Empfehlungen für Texte,

281 Das Beispiel der MAB-Daten habe ich entnommen: Kett, MABxml S. 1. Die Beispieldaten habe ich gekürzt (...).

282 <http://dublincore.org> (18.05.2008).

die mit Hilfe von *Auszeichnungssprachen* kodiert sind, zu etablieren (DC-Empfehlungen). Das standardisierte OpenDocument-Format²⁸³ greift beispielsweise auf DC-Vorschläge zur Kennzeichnung der Metadaten zurück.

Die *Dublin Core Metadata Initiative* schlägt in erster Linie Begriffe (terms) vor, die zur Bezeichnung der Daten, die eine elektronische Ressource beschreiben, verwendet werden sollen. Urheber sollen beispielsweise einheitlich als »creator« bezeichnet werden, Titel als »title« etc. Weitere normierte Begriffe betreffen die eindeutige Identifizierung von Quellen, Angaben zum Datenformat, zu Art (z. B. Text oder Image) sowie Sprache, Angaben zum Inhalt (Beschreibung, Schlagwörter) und zu Rechten an den Inhalten. Für einzelne Bereiche wird eine Standardisierung vorgeschlagen. So gilt die Empfehlung, als eindeutiges Identifizierungsmerkmal eine *URL* oder eine ISBN anzugeben. Bei der Angabe der Sprache wird empfohlen, sich an die Sprachkürzel nach ISO 639 zu halten und diese gegebenenfalls durch ein Länderkürzel nach ISO 3166 zu ergänzen.²⁸⁴

Da die DC-Empfehlungen im Wesentlichen eine Nomenklatur darstellen, mit deren Hilfe Angaben, die im Rahmen von Metadaten gemacht werden können, benannt werden sollen, können die Empfehlungen auf verschiedene Weise umgesetzt werden. So gibt es unterschiedliche Richtlinien, wie die DC-Empfehlungen in Verbindung mit *HTML* oder innerhalb von *XML* umgesetzt werden sollen.

Zur Veranschaulichung habe ich auf Angaben aus Abbildung 35 zurückgegriffen und diese in eine DC-konforme *XML*-Strukturform überführt.

```
<?xml version="1.0"?>
<metadata xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:format>Online-Ressource</dc:format>
  <dc:creator>Kafka, Franz</dc:creator>
  <dc:identifizier>
    http://erf.sbb.spk-berlin.de/han/381834964
  </dc:identifizier>
  <dc:title>Kafkas Werke im WWW</dc:title>
  <dc:description>
    Kritische Kafka-Ausgabe des S. Fischer Verlages bei Chadwyck-Healey.
    Parallele CD-ROM-Ausg. u.d.T.: Kafkas Werke
  </dc:description>
  <dc:publisher>
    ProQuest Information and Learning Co. Cambridge
  </dc:publisher>
  <dc:date>2004</dc:date>
  <dc:subject>
    18.10 Deutsche Literatur
    17.97 Texte eines einzelnen Autors
  </dc:subject>
</metadata>
```

Beispiel 18. Angaben aus Abbildung 35 nach Dublin Core kodiert

Anhand des Beispiels wird das Prinzip der Kodierung deutlich. Nach der XML-Deklaration und dem Root-Element »metadata«, in dem auch der *Namensraum* für die Angaben, die nach den DC-Empfehlungen ausgezeichneten Daten deklariert wird, folgt die Angabe »format«, die kennzeichnet, in welcher Form die betreffenden Daten vorliegen. Wie die Angabe auszusehen hat, wird durch die DC-Empfehlungen nicht genau spezifiziert. Das »creator«-Element kennzeichnet den Urheber des Textes. Eine Unterscheidung zwischen Autor und Herausgeber eines Werkes ist nicht vorgesehen. Gegebenenfalls kann »creator« mehrfach für Autoren und Herausgeber verwendet werden. Eine eindeutige Identifizierung der beschriebenen Ressource soll die Angaben des »identifizier« ermöglichen. Die DC-Empfehlungen schlagen unter anderem die Angabe des *URL* vor, unter der

283 Vgl.3.2.6 XML-Anwendungen, S. 91.

284 ISO 639 ist ein von der »Internationalen Organisation für Normung« festgelegter Standard, der Kennungen für Namen von Sprachen (Sprachcodes) definiert. Da einige Sprachen in verschiedenen Ländern gesprochen werden, kann zusätzlich ein Länderkürzel gemäß ISO 3166, einem Standard zur Kodierung geografischer Einheiten, ergänzt werden, um ggf. auf regionale Eigenheiten der Sprache hinzuweisen.

die Ressource aufgerufen werden kann. Das »title«-Element enthält den Titel des Werks, unter dem es formell bekannt ist. Der Untertitel und der Hinweis auf die parallele CD-ROM-Ausgabe ist unter »description« angegeben. Bei der Formulierung der Beschreibung lassen die DC-Empfehlungen alle Möglichkeiten offen. Der Verlag wird durch das »publisher« genannte Element gekennzeichnet. Hier sehen die DC-Empfehlungen vor, entweder eine verantwortliche Person oder eine Organisation anzugeben. Das Element »date« ist dazu vorgesehen, eine Zeitangabe in Bezug auf das Dokument anzugeben. Im obigen Beispiel kennzeichnet es das Publikationsjahr. Mit »subject« sollen schließlich Schlagwörter gekennzeichnet werden. Dieses Element bot sich zur Kennzeichnung der Sachgebiete an, denen das Werk im Katalog der Staatsbibliothek zugeordnet ist.

Da die DC-Empfehlungen nur an wenigen Stellen konkrete Vorgaben machen, wie bei der Erhebung und Formulierung von Metadaten vorzugehen ist, können sie lediglich als Rahmenwerk betrachtet werden. Die Formalisierung der Daten selbst muss auf einer anderen Ebene geregelt werden.

TEI-Header

Die *TEI*-Richtlinien sehen vor, deskriptive Metadaten im *TEI*-Header anzugeben. Auch wenn die *TEI*-Richtlinien hier diverse Elemente zur Kodierung entsprechender Informationen zur Verfügung stellen, gibt es kein verbindliches Regelwerk, wie die Daten genau zu erheben sind. Empfehlenswert wäre, auf ein geeignetes Regelwerk zurückzugreifen. Prinzipiell ist es z. B. möglich, die entsprechenden Daten nach RAK zu erheben. Allerdings lassen sich die komplexen Informationen einer Titelaufnahme nach RAK nicht komplett mit den von der *TEI* vorgeschlagenen Elementen kodieren.

Um die Möglichkeiten zu veranschaulichen, die die *TEI*-Richtlinien vorsehen, um Metadaten zu kodieren, greife ich auf die bibliografischen Angaben aus Abbildung 35 zurück. Die *TEI*-Richtlinien sehen eine Unterteilung der deskriptiven Metadaten in folgende Bereiche vor:

- Bibliografische Angaben (file description)
- Angaben zur Kodierung des Dokuments (encoding description)
- Beschreibung des Textprofils (text-profile description)
- Angaben zum Bearbeitungsstand (revision description)

Das Beispiel beschränkt sich auf die bibliografischen Angaben, was ausreicht, um das Prinzip einer Kodierung von Metadaten gemäß den *TEI*-Richtlinien zu verstehen.

```
<?xml version="1.0"??>
<TEI>
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Kafkas Werke im WWW</title>
        <author>Kafka, Franz</author>
      </titleStmt>
      <editionStmt>
        <edition>
          Kritische Kafka-Ausgabe des S. Fischer Verlages bei Chadwyck-
          Healey.
        </edition>
      </editionStmt>
      <publicationStmt>
        <publisher>ProQuest Information and Learning Co.</publisher>
        <pubPlace>Cambridge</pubPlace>
        <date>2004</date>
      </publicationStmt>
      <profileDesc>
        <classCode>18.10 Deutsche Literatur</classCode>
```

```
<classCode>17.97 Texte eines einzelnen Autors</classCode>
</profileDesc>
</fileDesc>
</teiHeader>
</TEI>
```

Beispiel 19. Angaben aus Abbildung 35 nach TEI kodiert

Die Möglichkeiten, die die *TEI*-Richtlinien zur Kodierung von Metadaten vorsehen, sind im Vergleich zu den DC-Empfehlungen sehr umfangreich und erlauben es, auch komplexe Metadaten zu beschreiben. Im Gegensatz zu den Kodierungsmodellen MAB oder DC ist das Konzept der Metadatenkodierung im *TEI*-Header nicht dazu ausgelegt, dass die Daten unabhängig vom Dokument verwendet werden. Entsprechend fehlen Möglichkeiten, das Datenformat anzugeben (es handelt sich zwangsläufig um Daten, die *TEI/XML*-kodiert sind) oder anzugeben, wo das Dokument zu finden ist (die Metadaten sind Teil des Dokuments).

Auf die Einzelheiten des Beispiels gehe ich nicht weiter ein, da es weitgehend selbst erklärend ist. Bemerkenswert ist, dass die *TEI*-Richtlinien neben den Möglichkeiten, Metadaten sehr genau zu qualifizieren, viele Bereiche berücksichtigen, die bei der Arbeit mit elektronischen Editionen relevant sind. So sind Angaben zu Details der *TEI*-Auszeichnung des Dokuments ebenso vorgesehen wie ausführliche Angaben zu den zugrunde liegenden Vorlagen und Quellen sowie zum Bearbeitungsstand der Daten.

Bis auf das MAB, das zu diesem Zweck auf die RAK zurückgreift, existiert zurzeit kein zuverlässiger Standard, nach dem bibliografische Daten formalisiert werden können, um sie anschließend in standardisierter Form im Rahmen elektronischer Editionen kodieren zu können. Den Richtlinien für den *TEI*-Header wie den Vorschlägen der *DCMI* mangelt es an einem zugrunde liegenden Regelwerk für eine spezialisierte einheitliche Erschließung bibliografischer Daten. Das MAB stellt in dieser Hinsicht das zurzeit komplexeste und etablierteste Format dar, ist aber im Sinne aktueller Anwendungen von *Auszeichnungssprachen* veraltet. Besonders im Hinblick auf die Erhebung semantischer Daten besteht hier der Bedarf nach einem modernen Standard, der auch weitere Aspekte, die über die bibliografischen Daten hinausgehen, nach einem einheitlichen Regelwerk erschließt.

4 Ausblick

Ausgangspunkt dieser Arbeit war die Beobachtung, dass bei der wissenschaftlichen Arbeit mit Texten offensichtlich nicht auf Papierausgaben verzichtet werden kann, obwohl sowohl die Erstellung von als auch die Arbeit mit Texten größtenteils mit Hilfe des Computers erfolgt.

Die Darstellung hat gezeigt, dass die Möglichkeiten bisheriger elektronischer Editionen nicht allen Anforderungen gerecht werden, die an eine solche Publikationsform zu stellen sind. Die bestehenden elektronischen Publikationskonzepte finden ihre Grenzen entweder bei der Darstellung und Präsentation von Texten, die nicht an die Möglichkeiten des Buchdrucks heranreichen, oder in den elektronischen Zugriffs- und Nutzungsmöglichkeiten der Texte. Einschränkungen bei der Textdarstellung bedeuten in den meisten Fällen einen Verlust an Informationen, die ein gedruckter Text enthalten kann und die für die wissenschaftliche Auseinandersetzung mit Texten relevant sein können; Einschränkungen bei den Möglichkeiten, Texte mit elektronischen Möglichkeiten und Technologien auszuwerten und weiterzuverarbeiten, stellen den Nutzen elektronischer Editionen generell in Frage.

Eine Perspektive im Umgang mit Texten in elektronischen Medien stellt die Verwendung von *Auszeichnungssprachen* wie die *XML* dar. Die Anwendung solcher *Auszeichnungssprachen* bei der Kodierung von Texten, wie sie beispielweise die *Text Encoding Initiative* vorschlägt, ermöglicht es einerseits, einen großen Teil der Informationen, die im Rahmen der wissenschaftlichen Arbeit mit Texten relevant sind, entweder im Text selber oder in zum Text gehörenden Metadaten zu bewahren. Andererseits bieten sie vielfältige Möglichkeiten, die Texte mit elektronischen Mitteln auszuwerten und zu kommentieren, bedarfsgerecht darzustellen und nicht zuletzt untereinander zu verlinken und zu vernetzen.

Auf dem Weg von der Gutenberg-Galaxis in ein digitales Dokuversum sind noch einige Etappen zu bewältigen.

Von Seiten der Germanistik fehlen bislang umfassende Untersuchungen über Informationsebenen und Informationen, die Texte auf Papier bzw. in Büchern enthalten. Es wäre wünschenswert, hier konkrete Beschreibungen und Benennungen zu entwickeln. Können solche Informationen explizit aufgeführt werden, lassen sie sich mit den vorgestellten Möglichkeiten von *Auszeichnungssprachen* kodiert in elektronische Texte einpflegen. Gegenstände solcher Untersuchungen wären etwa die semantischen Aspekte von Typografie und Layout, die Analyse von Dispositiven von Texten oder die Bedeutung der Distributionswege für die Rezeption von Büchern.

Die in dieser Arbeit aufgezeigten Möglichkeiten beim Umgang und bei der Arbeit mit Texten in elektronischen Medien setzt nicht zuletzt die Bereitschaft voraus, einen abstrakteren Umgang mit Texten zu akzeptieren. Die Konventionen, die sich im Umgang mit Texten auf Papier entwickelt haben, müssen thematisiert und reflektiert, Texte auf ihre medialen Bedingungen hin untersucht werden. Mit einem Textbegriff, der dem »papiernen« Denken verhaftet bleibt, wird man schwerlich die Möglichkeiten moderner Medien erschließen können.

Wünschenswert wäre darüber hinaus eine stärkere Beteiligung der geisteswissenschaftlichen Disziplinen bei der Entwicklung einheitlicher Standards zur Textauszeichnung bzw. Kodierung von Texten und Textinformationen mit Hilfe von *Auszeichnungssprachen*. Bei der Entwicklung geeigneter Werkzeuge für die Arbeit an *XML*-kodierten Texten sollten Erkenntnisse und Bedürfnisse aus dem Gebiet der Germanistik einfließen. Eine Einrichtung nach dem Vorbild der *Text Encoding Initiative*, die entsprechende Aktivitäten koordiniert und eine stärkere Beteiligung deutscher Forschungseinrichtungen an der *TEI* wären dabei wichtige Schritte.

Die Entwicklung von Computerprogrammen, die die Arbeit an und mit entsprechend kodierten Texten erleichtern, wäre eine Forderung, die an Programmierer und Informatiker zu richten ist. Die Voraussetzung hierfür ist, dass Standards vorgegeben werden, auf denen Programme zur Verarbeitung solcher Texte aufbauen können.

Glossar

American Standard Code for Information Interchange – ASCII

Der am weitesten verbreitete Computer-Zeichenkode. Er definiert 128 (7-Bit) bzw. 256 (8-Bit) Zeichen. Vgl. Grundlagen der Digitaltechnik, S. 71.

Auszeichnungssprache

Eine Auszeichnungssprache (Markup Language, ML) dient zur Beschreibung von Daten und legt die Regeln fest, nach denen die Beschreibung der Daten erfolgen soll. Ursprünglich dienten die Auszeichnungen im Text als Anweisungen für die Setzer im Drucksatz. Vgl. 3.1 Auszeichnungssprachen, S. 71.

Cascading Stylesheets – CSS

Eine Beschreibungssprache, mit deren Hilfe die Darstellung von Daten und Texten, die beispielsweise im *HTML*- und *XML*-Format vorliegen, innerhalb eines *Stylesheets* beschrieben werden kann. Vgl. Technologien zur Formatierung, S. 92.

DCMI

Abkürzung für die *Dublin Core Metadata Initiative*.

Dokument Type Definition – DTD

Ein Regelsatz zur Definition von *Inhaltsmodellen*. Vgl. DTD, S. 87.

Dublin Core Metadata Initiative

Unter dieser Bezeichnung haben sich verschiedene Institutionen zusammengeschlossen, um Konventionen zur Standardisierung von Metadaten, die Dokumente und andere Objekte im Internet einfacher auffindbar machen sollen, zu erarbeiten und zu etablieren. Vgl. Die Dublin Core Metadata Initiative, S. 133.

Extensible Hypertext Markup Language

Eine neue Version von *HTML*, die auf der Grundlage von *XML* definiert wurde.

Extensible Markup Language

(XML) Eine »erweiterbare« *Auszeichnungssprache*. Die zurzeit meistverwendete *Auszeichnungssprache*. Vgl. 3.1.4 Extensible Markup Language, S. 77.

Extensible Stylesheet Language

(XSL) Eine Familie von Sprachen, die zur Steuerung der Darstellung von *XML*-Dokumenten auf unterschiedlichen Ausgabemedien dient. Vgl. Technologien zur Formatierung, S. 92.

Font

Eine Schriftart bzw. Schriftfamilie zur Formatierung von Texten in elektronischen Medien. Während die Buchstaben im Computer in Form von Zeichenkodes wie *ASCII* oder *UNICODE* verarbeitet werden, wird das Aussehen der Buchstaben durch Fonts geregelt.

HTML, Hyper Text Markup Language, Hyperlinks, Hypermedia, Hypertext

Hyper Text Markup Language (*HTML*) ist eine *Auszeichnungssprache* die entwickelt wurde, um Texte und Informationen innerhalb des *World Wide Web* strukturieren und mit Hilfe der sogenannten Hyperlinks vernetzen zu können. Ein System entsprechend vernetzter Informationen bezeichnet man als Hypermedia. Vgl. 1.4.3 Hypertext und Hypermedia, S. 22 und 3.1.3 Hyper Text Markup Language, S. 76.

Inhaltsmodell

Ein Inhaltsmodell dient zur Definition von Elementen, Attributen, Entitäten und Inhalten, die bei der Anwendung einer *Auszeichnungssprache* wie XML oder SGML erlaubt sind. Inhaltsmodelle werden mit Hilfe spezieller Regelsätze wie DTD, XML-Schema oder RELAX-NG formuliert. Vgl. 3.2.4 XML-Inhaltsmodelle, S. 87.

Namensräume, Namespaces

Namensräume erlauben es, innerhalb eines XML-Dokuments Elemente aus unterschiedlichen XML-Anwendungen zu verwenden und diese zu unterscheiden, zu welcher Anwendung ein verwendetes Element gehört. Vgl. 3.2.5 XML-Namensräume, S. 88.

Portable Document Format – PDF

Ein plattformübergreifendes Dateiformat für Dokumente, das von der Firma Adobe Systems entwickelt wurde. Es erlaubt die einheitliche Darstellung von Dokumenten auf unterschiedlichen Computersystemen und Ausgabemedien. Vgl. 2.2.4 Datenformate für formatierte Texte, S. 60.

PostScript

Eine Programmiersprache zur Darstellungssteuerung bei der Ausgabe elektronisch erstellter Dokumente, die von der Firma Adobe entwickelt wurde. PostScript hat sich einem Standard in der Druckindustrie entwickelt.

Regular Language Description for XML New Generation – RELAX-NG

Ein Regelsatz zur Definition von *Inhaltsmodellen*. Vgl. Relax NG, S. 88.

Retrieval

Die Anwendung computergestützter inhaltsorientierter Suchmethoden. Wie der Begriff Retrieval (deutsch: Wiedergewinnung, Auffindung) sagt, sind Informationen in großen Datenbeständen zunächst verloren und müssen wieder gewonnen bzw. wieder gefunden werden. Vgl. 1.4.4 Retrieval, S. 23.

Standard Generalized Markup Language – SGML

Eine standardisierte *Auszeichnungssprache*. Vgl. 3.1.2 Standard Generalized Markup Language, S. 74.

Stylesheet

Ein Satz von Regeln, der angibt, wie Texte, die in einer *Auszeichnungssprache* vorliegen, dargestellt werden – ähnlich einer Formatvorlage. Vgl. Technologien zur Formatierung, S. 92.

Tag

Eine Auszeichnungsmarkierung, wie sie *Auszeichnungssprachen* verwenden. Vgl. Tags, S. 82.

Text Encoding Initiative – TEI

Eine 1987 gegründete Organisation und ein gleichnamiges Dokumentenformat zur Kodierung von Texten auf der Basis von SGML/XML. TEI hat sich zu einem De-facto-Standard innerhalb der Geisteswissenschaften entwickelt. Vgl. 3.3 Die Text Encoding Initiative, S. 94.

UNICODE

Ein internationaler Zeichenkode, in dem für jedes Schriftzeichen bzw. Textelement aller bekannten Schriftkulturen und Zeichensysteme ein digitaler Kode festgelegt wird. Im Gegensatz zu ASCII, können mit Hilfe der 16-Bit Datenwörter, die der Unicode verwendet, 65536 Zeichen kodiert werden. Vgl. Grundlagen der Digitaltechnik, S. 71.

Uniform Resource Locator – URL

Die »Internetadresse«, unter der Daten, vor allem im WWW, zu finden sind.

World Wide Web – WWW

Auch W3. Ein über das Internet abrufbares *Hypertext*-System. Mit Hilfe von Webbrowsern können die Daten von Webservern abgerufen werden. Der Benutzer kann den Hyperlinks im Dokument folgen, die auf andere Dokumente im *WWW* verweisen. Hierdurch ergibt sich ein weltweites Netz (oder Gewebe) aus Webseiten.

XHTML

Abkürzung für *Extensible Hypertext Markup Language*.

XLink

Abkürzung für *XML Linking Language*.

XML

Abkürzung für *Extensible Markup Language*.

XML Path Language

Eine Anfragesprache, um Teile eines *XML*-Dokumentes zu adressieren. *XPath* dient als Grundlage einer Reihe weiterer Standards wie *XSLT* und *XPointer*. Vgl. Technologien zur Adressierung, S. 93.

XML Pointer Language

Eine Anfragesprache, um Teile eines *XML*-Dokumentes zu adressieren. Es handelt sich dabei um eine Erweiterung der *XML Path Language*-Spezifikation. Vgl. Technologien zur Adressierung, S. 93.

XML-Anwendung

Auszeichnungssprachen wie *XML* stellen lediglich eine »Grammatik« zur Verfügung, die angibt, in welcher Form die Auszeichnung von Daten zu erfolgen hat. Für jede Anwendung von *XML* muss ein konkretes »Vokabular« festgelegt werden, das mit Hilfe eines Inhaltsmodells definiert werden kann. Vgl. 3.2.6 *XML*-Anwendungen, S. 91.

XML Linking Language

Eine Syntax zur Definition von Links in *XML*-Dokumenten. Dabei können Verbindungen von einem Punkt A zu einem Punkt B angegeben werden (ähnlich den aus *HTML* bekannten *Hyperlinks*). Es können jedoch auch Dokumente in zwei Richtungen (von A nach B und zurück) sowie multidirektional (zwischen beliebig vielen *XML*-Dokumenten) verlinkt werden. Vgl. Technologien zur Adressierung, S. 93.

XML-Schema

Ein Regelsatz zur Definition von *Inhaltsmodellen*.

XPath

Abkürzung für *XML Path Language*.

XPointer

Abkürzung für *XML Pointer Language*.

XSL

Abkürzung für *Extensible Stylesheet Language*.

Literatur

Quellen

- Adelung, Johann Christoph: Grammatisch-kritisches Wörterbuch der hochdeutschen Mundart: mit beständiger Vergleichung der übrigen Mundarten, besonders aber der oberdeutschen. Elektronische Volltext- und Faks.-Ed. nach der Ausg. letzter Hand, Leipzig 1793–1801. Berlin: Directmedia Publ., 2001.
- Adelung, Johann Christoph: Grammatisch-kritisches Wörterbuch der Hochdeutschen Mundart: mit beständiger Vergleichung der übrigen Mundarten, besonders aber der Oberdeutschen. Mit D. W. Soltau's Beyträgen revidirt und berichtigt von Franz Xaver Schönberger. Wien: Bauer, 1811. <http://mdz.bib-bvb.de/digbib/lexika/adelung> (18.05.2008).
- Bertram, Mathias (Hg.): Deutsche Literatur von Lessing bis Kafka. Studienbibliothek. Berlin: Directmedia Publishing, 2000.
- Büchner, Georg: Sämtliche Werke und Schriften. Historisch-kritische Ausgabe mit Quellendokumentation und Kommentar. Hg. von Burghard Dedner und Thomas Michael Mayer. 14 Bde. Darmstadt: Wissenschaftliche Buchgesellschaft, 2000 ff.
- Burch, Thomas u. a. (Hg.): Mittelhochdeutsche Wörterbücher im Verbund. CD-ROM und Begleitbuch. Stuttgart: Hirzel, 2002.
- Busa, Roberto: Index Thomisticus. Sancti Thomae Aquinatis operum omnium indices et concordantiae in quibus verborum omnium et singulorum formae et lemmata cum suis frequentiiis et contextibus variis modis referuntur, quaeque auspice Paulo VI Summo Pontifice consociata plurimum opera atque electronico IBM automato usus digessit Robertus Busa. Stuttgart, Bad Cannstadt: Frommann-Holzboog, 1974–80.
- Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm. Elektronische Ausgabe der Erstbearbeitung. Hg. vom Kompetenzzentrum für Elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier... Bearb. von Hans-Werner Bartz... Frankfurt a. M.: Zweitausendeins, 2004.
- Die Fackel. Volltextausgabe. Hg von Karl Kraus. Mit Bibliographie und Register. Hg von Wolfgang Hink. München: Saur; Frankfurt a. M.: Zweitausendeins, 2002. CD-ROM.
- Goethe, Johann Wolfgang von: Der junge Goethe in seiner Zeit. Texte und Kontexte. Hg von Karl Eibl u. a. Frankfurt a. M.: Insel, 1998. 2 Bde. und 1 CD-ROM.
- Goethe, Johann Wolfgang von: Werke: Hg. von Mathias Bertram. CD-ROM. Berlin: Directmedia 1998.
- Goethe, Johann Wolfgang von: Zeit, Leben, Werk. Von Jürgen von Esenwein/Harald Gerlach. Stuttgart u. a.: Metzler u. a. 1999. CD-ROM.
- Goethes Werk im Kontext. Die »Berliner Ausgabe« auf CD-ROM. Berlin: Carsten Worm InfoSoftWare, 2004.
- Goethes Werke auf CD-ROM. Weimarer Ausgabe. Cambridge: ProQuest Information and Learning Co., 1997-
- Goethes Werke im WWW. Weimarer Ausgabe. Cambridge: ProQuest Information and Learning Co., 1997-
- Gotthold Ephraim Lessings sämtliche Schriften. Hg. von Karl Lachmann. Berlin, 1938 ff.
- Harbou, Thea von: Metropolis. Drehbuch-Typoskript. Im Besitz der Stiftung Deutsche Kinemathek, Berlin.
- Kafka, Franz. Kritische Ausgabe. Schriften, Tagebücher, Briefe. Hg. von Jürgen Born u. a. Frankfurt a. M.: Fischer, 1982 ff.
- ders.: Historisch-kritische Ausgabe sämtlicher Handschriften, Drucke und Typoskripte. Hg. von Roland Reuß. Basel/Frankfurt a. M.: Stroemfeld, 1995 ff.
- ders.: Kafkas Werke im WWW/Kafkas Werke auf CD-ROM. Kritische Kafka-Ausgabe des S. Fischer Verlages bei Chadwyck-Healey. Chadwyck-Healey, 1999. <http://kafka.chadwyck.com> (18.05.2008).
- Keller, Gottfried. Sämtliche Werke. Hg. von Walter Morgenthaler. Basel u. a.: Stroemfeld u. a., 1996 ff.
- Kindlers neues Literatur-Lexikon. Das 23-bändige Werk auf einer CD-ROM. München: Systema u. a., 2000.
- Kritisches Lexikon zur deutschsprachigen Gegenwartsliteratur. Hg. von Heinz Ludwig Arnold. München: Edition Text Kritik. Als Loseblattsammlung, auf CD-ROM und online. <http://www.klgonline.de> (18.05.2008).

- Lessing, Gotthold Ephraim: Die alte Jungfer. Ein Lustspiel in drey Aufzügen. Berlin: 1749. Wiedergabe in: Bibliothek der deutschen Literatur. Eine Ed. der Kulturstiftung der Länder. Mikrofiche-Gesamtausg. nach den Angaben des Taschengoedeke. München u. a.: Saur, [1994]. Fich 19867.
- ders.: Gesammelte Werke. Berlin: Aufbau, 1954.
- ders.: Gotthold Ephraim Lessings sämtliche Schriften. Hg. von Karl Lachmann. 3., auf's neue durchges. und verm. Aufl. Besorgt durch Franz Muncker. Stuttgart: Göschen, 1887.
- ders.: Lessings Werke. Hg. von Franz Boxmüller. Leipzig/Wien: Bibliographisches Institut, [1908].
- ders.: Werke in drei Bänden. München: Artemis u. Winkler, 31994.
- ders.: Werke. Herausgegeben von Herbert G. Göpfert. 8 Bde. München: Hanser, 1970 ff.
- Morgenstern, Christian: Alle Galgenlieder. Zürich: Diogenes, 1981.
- Musil, Robert: Robert Musil. Der literarische Nachlaß. Hg. von Friedbert Aspetsberger u. a. Elektronisch ediert (Wordcruncher) von Karl Eibl und Marianne Willems. Reinbek: Rowohlt, 1992.
- Zedler, Johann Heinrich: Grosses vollständiges Universal-Lexicon aller Wissenschaften und Künste. Leipzig 1732–1754. <http://mdz10.bib-bvb.de/~zedler/zedler2007/index.html> (18.05.2008).

Forschungsliteratur

- Allers, Heinrich und Nickel, Gunter: Die EDV-gestützte Erstellung kommentierter Register mit allegro-C. In: editio. Internationales Jahrbuch für Editionswissenschaft 10 (1996). S. 140-151.
- Altrichter, Helmut: Retrodigitalisierung in Deutschland – Versuch einer Zwischenbilanz. <http://www.bsb-muenchen.de/mdz/forum/altrichter/> (18.05.2008).
- Arnold, Heinz Ludwig und Detering, Heinrich (Hg.): Grundzüge der Literaturwissenschaft. München: dtv, 2002.
- Assmann, Aleida: Aspekte einer Materialgeschichte des Lesens. In: Gestern begann die Zukunft. Entwicklung der gesellschaftlichen Bedeutung der Medienvielfalt. Hg. von Hilmar Hoffmann. Darmstadt: WBG, 1994. S. 3-16.
- Assmann, Bernhard: Heinrich IV. und XML. In: Fundus – Forum für Geschichte und ihre Quellen. Heft 2. <http://webdoc.gwdg.de/edoc/p/fundus/2/assmann.pdf> (18.05.2008).
- Aumann, Stefan u. a.: From Digital Archive to Digital Edition. In: Historical Social Research 1/24 (1999). S. 101-144.
- Baasner, Rainer: Digitalisierung – Geisteswissenschaften – Medienwechsel? Hypertext als fachgerechte Publikationsform. In: Jahrbuch für Computerphilologie 1 (1999). S. 11-20.
- Bader, Winfried: Erstellen einfacher Register mit TUSTEP. In: Historical Social Research 15 (1990). Nr. 2, S. 119-137.
- ders.: Was ist die Text Encoding Initiative (TEI)? In: Computergestützte Text-Edition. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu Editio 12). S. 9-20.
- Barckow, Klaus und Delabar, Walter (Hg.): Neue Informations- und Speichermedien in der Germanistik. Zu den Perspektiven der EDV als Informationsträger für die literaturwissenschaftliche Forschung. Bern u. a.: Lang, 1994.
- Bartz, Hans-Werner u. a.: Beyond the Scope of the Printed Dictionary. Digitizing the »Deutsche Wörterbuch« by Jacob and Wilhelm Grimm. In: Standards and Methoden der Volltextdigitalisierung. Beiträge des internationalen Kolloquiums an der Universität Trier. 8./9. Oktober 2001. Akademie der Wissenschaften und der Literatur Mainz. Stuttgart: Steiner, 2003. S. 291-295.
- Baßler, Moritz: Stichwort Text. Die Literaturwissenschaft unterwegs zu ihrem Gegenstand. In: Jahrbuch der Deutschen Schiller-Gesellschaft 42 (1998). S. 470-475.
- Bock, Wolfgang: Bild – Schrift – Cyberspace. Grundkurs Medienwissen. Bielefeld: Aisthesis, 2002.
- Bohnenkamp, Anne: Textkritik und Textedition. In: Grundzüge der Literaturwissenschaft. Hg. von Heinz Ludwig Arnold und Heinrich Detering. München: dtv, 1997. S. 179-203.
- Bolz, Norbert u. a. (Hg.): Computer als Medium. München: Fink, 1994.

- Brahms, Ewald: Auf dem Weg zu einer verteilten Digitalen Forschungsbibliothek. In: Fundus – Forum für Geschichte und ihre Quellen. Heft 5. <http://webdoc.gwdg.de/edoc/p/fundus/5/brahms.pdf> (18.05.2008).
- Brunkmann, Marlies und Susen, Gerd-Hermann: Erschließungsaspekte elektronischer Briefeditionen. In: Edition und Internet. Hg. von Jörg Meier und Arne Ziegler. Berlin: Weidler, 2004. (Beiträge zur Editionsphilologie 2). S. 79-95.
- Büchel, Gregor und Schröder, Bernhard: Verfahren und Techniken in der computergestützten Lexikographie. In: Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. Hg. von Ingrid Lemberg u. a. Tübingen: Niemeyer, 2001. (Lexicographica: Series maior 107). S. 7-28.
- Burch, Thomas u. a. (Hg.): Standards und Methoden der Volltextdigitalisierung. Beiträge des internationalen Kolloquiums an der Universität Trier. 8./9. Oktober 2001. Akademie der Wissenschaften und der Literatur Mainz. Stuttgart: Steiner, 2003.
- ders.: Arbeiten des Kompetenzzentrums für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier. In: Jahrbuch der Akademie der Wissenschaften und der Literatur 52 (2001). S. 251-264.
- ders.: Ein »Hausbuch« für alle? – Das Deutsche Wörterbuch von Jacob und Wilhelm Grimm auf CD-ROM und im Internet. In: Jahrbuch für Computerphilologie 2 (2000). S. 11-34.
- dies.: Mittelhochdeutsche Wörterbücher auf CD-ROM und im Internet. Der Einsatz von SGML in der Retrodigitalisierung lexikographischer Standardwerke. In: Akademie-Journal. Mitteilungsblatt der Konferenz der deutschen Akademien der Wissenschaften 2/1998. S. 17-24.
- ders. und Fournier, Johannes: Vom Buch zur elektronischen Publikation – Textdigitalisierung auf der Basis von SGML/XML. In: Automatische Textanalyse. Systeme und Methoden zur Annotation und Analyse natürlichsprachiger Texte. Hg. von Alexander Mehler und Henning Lobin. Wiesbaden: VS Verlag, 2004. S. 265-283.
- dies.: Zur Anwendung der TEI-Richtlinien auf die Retrodigitalisierung mittelhochdeutscher Wörterbücher. Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. Hg. von Ingrid Lemberg u. a. Tübingen: Niemeyer, 2001. (Lexicographica: Series maior 107). S. 133-153.
- Burnard, Lou und Bauman, Syd: TEI P5: Guidelines for Electronic Text Encoding and Interchange. TEI Consortium 2007. <http://www.tei-c.org/Guidelines/P5/> (18.05.2008).
- Bush, Vannevar: As We May Think. In: The Atlantic Monthly 7/1945 (Vol. 176, Heft 1). S. 101-108.
- Castrillo Benito, Nicolàs und Stahl, Peter (Hg.): TUSTEP educa. Actas de Congreso del International TUSTEP User Group. Penaranda de Duero (Burgos) Octubre 1999. Burgos: Univ., 2001.
- Charlier, Robert: Goethe Lexikographie im Internet. Aspekte prospektiver Digitalisierung. In: Edieren in der elektronischen Ära. Hg. von Gottfried Reeg, und Martin J. Schubert. Berlin: Weidler, 2004. S. 17-37.
- Chartier, Roger: Lesewelten. Buch und Lektüre in der frühen Neuzeit. Frankfurt a. M./New Yorck: Campus, 1990.
- Chenu, Marie-Dominique: Das Werk des hl. Thomas von Aquin. Verz. u. Erg. d. Arbeitshinweise von Otto M. Pesch. Vom Verf. Durchges. u. verb. dt. Ausg. 2. Aufl. Graz u. a.: Styria, 1982.
- Christmann, Ruth u. a.: Ein »heiligthum der sprache« digitalisiert: Das Deutsche Wörterbuch von Jacob und Wilhelm Grimm auf CD-ROM und im Internet. In: TUSTEP educa. Actas de Congreso del International TUSTEP User Group. Penaranda de Duero (Burgos) Octubre 1999. Hg. von Nicolàs Benito Castrillo und Peter Stahl. Burgos: Univ., 2001. S. 13-37.
- dies.: Books into Bytes: Jacob and Wilhelm Grimm's Deutsches Wörterbuch on CD-ROM and on the Internet. In: Literary and Linguistic Computing 1 (2001). S. 121-133.
- Computers and the humanities. Official journal of the Association for Computers and the Humanities. Special issue on Electronic publishing 32 (1998). (H. 4, S. 254-322).
- Coy, Wolfgang: Aus der Vorgeschichte des Mediums Computer. In: Computer als Medium. Hg. von Norbert, Bolz, Friedrich A. Kittler, Christoph Tholen. München: Fink, 1994. S. 19-37.

- Cyganiak, Richard: Publishing mit XML: Open eBook und DocBook. Ausarbeitung im Rahmen des Seminars XML-Technologien von Prof. Dr. Robert Tolksdorf, Freie Universität Berlin, Institut für Informatik, 2003.
- Delabar, Walter: Community. Über die Implementierung literaturhistorischer Informationen ins World Wide Web. In: Edition und Internet. Hg. von Jörg Meier und Arne Ziegler. Berlin: Weidler, 2004. (Beiträge zur Editionsphilologie 2). S. 97-110.
- Dieter, Jörg: Historisch-Kritische Edition im virtuellen Raum. Möglichkeiten und Grenzen elektronischer Editionen vor dem Hintergrund einiger Kernprobleme der Editionswissenschaft. 2002. <http://www.jolifanto.de/wissenschaft/webrhetorik/Textedition.pdf> (18.05.2008).
- Driscoll, Adrian und Scott, Brad: Electronic Publishing at Routledge. An Overview and Case Studies. In: Computers and the humanities. Official journal of the Association for Computers and the Humanities. Special issue on Electronic publishing. Jg. 32 (1998). S. 257-270.
- Eco, Umberto: Über Gott und die Welt. Essays und Glossen. München [u. a.]: Hanser, 1986.
- EDV-Projekte auf dem Gebiet der deutschen und altnordischen Sprache und Literatur. Bericht 5: Stand vom Herbst 1992. In: Germanistik 34(1993). H. 1. S. 427-430.
- EDV-Projekte auf dem Gebiet der deutschen und altnordischen Sprache und Literatur. Bericht 6: Stand vom Herbst 1993. In: Germanistik 35(1994). H. 1. S. 345-349.
- EDV-Projekte auf dem Gebiet der deutschen und altnordischen Sprache und Literatur. Bericht 7: Stand vom Herbst 1994. In: Germanistik 36. (1995). H. 1. S. 334-338.
- EDV-Projekte auf dem Gebiet der deutschen und altnordischen Sprache und Literatur. Bericht 8: Stand vom Herbst 1995. In: Germanistik 37 (1996). H. 1. S. 356-360.
- Eibl, Karl u. a.: Der junge Goethe in neuer Ausgabe. Einige Präliminarien und Marginalien. In: Computergestützte Text-Edition. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu Editio 12). S. 69-78.
- Eisenstein, Elisabeth L.: Die Druckerpresse. Kulturrevolutionen im frühen modernen Europa. Wien/New York: Springer, 1997.
- Eremen, Harald: Suchen, finden, lesen. WordCruncher – Text- und Retrieval Software. In: c't 3/1991. S. 152 ff.
- Eversberg, Bernhard: Was sind und was sollen Bibliothekarische Datenformate. Überarbeitete und erweiterte Neuauflage. Braunschweig: Univ.-Bibl., 1994. WWW-Version mit Ergänzungen 1999. <http://www.allegro-c.de/formate/formate.htm> (18.05.2008).
- Fabian, Bernhard und Kranz, Dieter: Interne Kollation. Eine Einführung in die maschinelle Textverglei- chung. In: Texte und Varianten. Probleme ihrer Edition und Interpretation. Hg. von Gunter Martens und Hans Zeller. München: Beck, 1971. S. 385-400.
- Fanta, Walter: Die Computer-Edition des Musil-Nachlasses. Baustein einer Epochendatenbank der Moderne. In: editio. Internationales Jahrbuch für Editions- wissenschaft 8 (1994). S. 127-157.
- Faulstich, Werner: Medienwissenschaft. Paderborn: Fink, 2004.
- Fetz, Bernhard und Kastberger, Klaus (Hg.): Von der ersten zur letzten Hand. Theorie und Praxis der literarischen Edition. Wien/Bozen: Folio, 2000.
- Finneran, Richard (Hg.): The Literary Text in the Digital Age. Michigan: University of Michigan Press, 1996.
- Firchow, Evelyn S.: Anforderungsprofil an eine Edition mittelalterlicher Werke, dargestellt am Beispiel des altisländischen »Elucidarius«. In: Historische Edition und Computer. Möglichkeiten und Probleme interdisziplinärer Textverarbeitung und Textbearbeitung. Hg. von Anton Schwob u. a. Graz: Leykam, 1989. S. 29-43.
- Flügge, Lars: Die Auswirkungen des Buchdrucks auf die Praxis des Schreibens. Marburg: Tectum, 2005.
- Fournier, Johannes: Digitale Dialektik. Chancen und Probleme mittelhochdeutscher Wörterbücher in elektronischer Form. In: Wörterbücher in der Diskussion IV. Vorträge aus dem Heidelberger Lexikographischen Kolloquium. Hg. von Herbert Ernst Wiegand. Tübingen: Niemeyer, 2000. S. 85-108.
- ders.: Mittelhochdeutsche Wörterbücher digital: Konzepte – Methoden – Entwicklung. <http://193.174.98.10/scan1/MDZ/kolloquium/ref/fo> (08.05.2000).

- ders.: Vom Datenstrom zur Datenhierarchie. TUSTEP als Werkzeug zur strukturierten Erfassung linearer Texte. In: TUSTEP educa. Actas de Congreso del International TUSTEP User Group. Peñaranda de Duero (Burgos). Octubre 1999. Hg. von Nicolàs Castrillo Benito und Peter Stahl. Burgos: Univ., 2001. S. 37-59.
- Frank, Barbara: Die Textgestalt als Zeichen. Lateinische Handschriftentradition und die Verschriftlichung der romanischen Tübingen: Narr, 1994.
- Funke, Fritz: Buchkunde. Die historische Entwicklung des Buches von der Keilschrift bis zur Gegenwart. 6. überarb. und erg. Aufl. Wiesbaden: Albus, 2006.
- Gabler, Hans Walter und Kreitmair, Wolfgang: Der Computer als Arbeitshilfe für das wissenschaftliche Edieren. In: Computerfibel für die Geisteswissenschaften. Einsatzmöglichkeiten des Personal Computers und Beispiele aus der Praxis. Hg. von Bernd Gregor und Manfred Krifka. München: Beck, 1987. S. 203-211.
- Gärtner, Kurt und Krummacker, Hans-Henrik (Hg.): Zur Überlieferung, Kritik und Edition alter und neuerer Texte: Beiträge des Colloquiums zum 85. Geburtstag von Werner Schröder am 12. und 13. März 1999 in Mainz. Stuttgart: Steiner, 2000.
- Giesecke, Michael: Der Buchdruck in der frühen Neuzeit. Eine historische Fallstudie über die Durchsetzung neuer Informations- und Kommunikationstechnologien. Frankfurt a. M.: Suhrkamp, 1991.
- ders.: Sinnenwandel, Sprachwandel, Kulturwandel. Studien zur Vorgeschichte der Informationsgesellschaft. Frankfurt a. M.: Suhrkamp, 1992.
- Gloning, Thomas und Welter, Rüdiger: Wortschatzarchitektur und elektronische Wörterbücher: Goethes Wortschatz und das Goethe-Wörterbuch. In: Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. Hg. von Ingrid Lemberg u. a. Tübingen: Niemeyer, 2001. (Lexicographica: Series maior 107). S. 117-132.
- Göttsche, Dirk: Ausgabentypen und Ausgabenbenutzer. In: Text und Edition. Positionen und Perspektiven. Hg. von Rüdiger Nutt-Kofoth. Berlin: Schmidt, 2000. S. 37-63.
- Graber, Stefan: Der Autorentext in der historisch-kritischen Ausgabe. Ansätze zu einer Theorie der Textkritik. Bern u. a.: Lang, 1998.
- Greenstein, Daniel: Publishing scholarly information in a digital millennium. In: Computers and the humanities. Official journal of the Association for Computers and the Humanities. Special issue on Electronic publishing. Jg. 32 (1998). S. 253-256.
- Gregor, Bernd und Krifka, Manfred (Hg.): Computerfibel für die Geisteswissenschaften. Einsatzmöglichkeiten des Personal Computers und Beispiele aus der Praxis. München: Beck, 1987.
- ders.: Sonderzeichen-Verarbeitung. In: Computerfibel für die Geisteswissenschaften. Einsatzmöglichkeiten des Personal Computers und Beispiele aus der Praxis. Hg. von Bernd Gregor und Manfred Krifka. München: Beck, 1987. S. 127-147.
- ders.: Vom PC zur Fotosatzanlage. In: Computerfibel für die Geisteswissenschaften. Einsatzmöglichkeiten des Personal Computers und Beispiele aus der Praxis. Hg. von Bernd Gregor und Manfred Krifka. München: Beck, 1987. S. 148-158.
- Gronemeyer, Horst: Der Philologe und sein Text in Handschrift, Buch und Datenbank. Vorgelegt in der Sitzung vom 18. Januar 2002. Göttingen: Vandenhoeck & Ruprecht u. a., 2002.
- Gumbrecht, Hans U. und Pfeiffer, K. Ludwik (Hg.): Materialität der Kommunikation. Frankfurt a. M.: Suhrkamp, 1995.
- Gumpert, Johann Peter: Zur »Typographie« der geschriebenen Seite. In: Pragmatische Schriftlichkeit im Mittelalter. Erscheinungsformen und Entwicklungsstufen. Hg. von Hagen Keller u. a. München: Fink, 1992. S. 283-292.
- Günzburger, Angelika: Fragen des Einsatzes von EDV bei der Erschließung von Nachlässen und Autographen. Kolloquium der Deutschen Forschungsgemeinschaft im Deutschen Literaturarchiv in Marbach a. N. am 25./26. Mai 1987. In: editio. Internationales Jahrbuch für Editionswissenschaft 2 (1988). S. 217-224.
- Hall, Steven: Literature Online – building a Home for English and American Literature on the World Wide Web. In: Computers and the humanities. Official journal of the Association for Computers and the Humanities. Special issue on Electronic publishing 32 (1998). S. 285-301.

- Härtel, Reinhard (Hg.): Geschichte und ihre Quellen. Festschrift für Friedrich Hausmann zum 70. Geb. Graz: Akad. Dr. u. Verl. Anst., 1987.
- ders.: Mehr als ein Anhang: Das computererstellte Register. In: Historische Edition und Computer. Möglichkeiten und Probleme interdisziplinärer Textverarbeitung und Textbearbeitung. Hg. von Anton Schwob u. a. Graz: Leykam, 1989. S. 67-84.
- Haß-Zumkehr, Angelika: Zur Mikrostruktur im Hypertext-Wörterbuch. In: Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. Hg. von Ingrid Lemberg u. a. Tübingen: Niemeyer, 2001. (Lexicographica: Series maior 107). S. 103-115.
- Heibach, Christiane: Literatur im Internet. Theorie und Praxis einer kooperativen Ästhetik. Berlin: dissertation.de, 2000.
- Henkes, Christiane (Hg.): Schrift – Text – Edition. Hans Walter Gabler zum 65. Geburtstag. Tübingen: Niemeyer, 2003. (Beihefte zu Editio 19).
- Henze, Volker: Metadata/Dublin Core. In: Computergestützte Text-Edition. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu Editio 12). S. 29-40.
- Hildenbrandt, Vera und Schares, Thomas: Das Grimmsche Wörterbuch geht ins 21. Jahrhundert: Präsentation eines Prototyps des digitalen Deutschen Wörterbuchs von Jacob und Wilhelm Grimm. <http://www.bsb-muenchen.de/mdz/forum/hildenbrandt/index.htm> (18.05.2008)
- Hille, Gunter: Projekt Gutenberg-DE. In: Computergestützte Text-Edition. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu Editio 12). S. 43-46.
- Hoffmann, Dirk u. a.: Computer-Edition statt Buch-Edition. Notizen zu einer historisch-kritischen Edition – basierend auf dem Konzept von hypertext und hypermedia. In: editio. Internationales Jahrbuch für Editions-wissenschaft 7 (1993). S. 211-220.
- Hoffmann, Hilmar (Hg.): Gestern begann die Zukunft. Entwicklung der gesellschaftlichen Bedeutung der Medienvielfalt. Darmstadt: WBG, 1994.
- Hurlebusch, Klaus: Kritische Edition als Datenverarbeitung? Eine neue Ausgabe der »Familie Ghonorez« und der »Familie Schroffenstein«. In: Kleist-Jahrbuch 1995. S. 183-199.
- Idensen, Heiko; Matthias Krohn: Bild-Schirm-Denken. Manual für hypermediale Diskurstechniken. In: Computer als Medium. Hg. von Norbert, Bolz, Friedrich A. Kittler, Christoph Tholen. München: Fink, 1994. S. 245-266.
- Illich, Ivan: Im Weinberg des Textes. Als das Schriftbild der Moderne entstand. Ein Kommentar zu Hugos »Didascalicon«. Frankfurt a. M.: Luchterhand, 1991.
- Jannidis, Fotis: Computerphilologie. Stuttgart: Metzler, 2007.
- ders.: Das Projekt <http://computerphilologie.uni-muenchen.de> im Kontext. In: Computergestützte Text-Edition. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu Editio 12). S. 61-68.
- ders.: TEI in der Praxis. Arbeitsbericht. <http://computerphilologie.tu-darmstadt.de/praxis/teiprax.html> (18.05.2008).
- ders.: Wider das Altern elektronischer Texte: philologische Textauszeichnung mit TEI. In: editio. Internationales Jahrbuch für Editions-wissenschaft 11 (1997). S. 152-177.
- Jenks, Stuart und Marra, Stephanie (Hg.): Internet-Handbuch Geschichte. Köln u. a.: Böhlau, 2001.
- Kammer, Manfred: Literarische Datenbanken: Anwendungen der Datenbanktechnologie in der Literaturwissenschaft. München: Fink, 1995.
- Kammer, Stephan: Jenseits des »Texts«. Zwei neue Studien zum Schreiben und zur Typographie. In: Text. Kritische Beiträge 8 (2003), S. 119-125.
- Kamphusmann, Thomas: Literatur auf dem Rechner. Stuttgart: Metzler, 2002.
- Kamzelak, Roland (Hg.): Computergestützte Text-Edition. Tübingen: Niemeyer, 1999. (Beihefte zu Editio 12.)
- ders.: Edition und EDV. Neue Editions-praxis durch Hypertext-Editionen. In: Text und Edition. Positionen und Perspektiven. Hg. von Rüdiger Nutt-Kofoth. Berlin: Schmidt, 2000. S. 65-80.

- ders.: Hypermedia – brauchen wir eine neue Editionswissenschaft? In: Computergestützte Text-Edition. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu Editio 12). S. 117-126.
- ders.: Hypermedia und Philologie. In: Jahrbuch für Computerphilologie 1 (1999), S. 61-70.
- ders.: Wider ein Verfallsdatum von Editionsdaten. Propositum. In: Edition und Internet. Hg. von Jörg Meier und Arne Ziegler. Berlin: Weidler, 2004. (Beiträge zur Editionsphilologie 2). S. 67-78.
- Kanzog, Klaus: Einführung in die Editionsphilologie der neueren deutschen Literatur. Berlin: Schmidt, 1991.
- ders.: Fixierter Text – realisierter Text. Über eine vernachlässigte Aufgabe der Editionsphilologie. In: Edition als Wissenschaft. Festschrift für Hans Zeller. Hg. von Gunter Martens und Winfried Woesler. Tübingen: Niemeyer, 1991. (Beihefte zu Editio 2). S. 5-16.
- Kapp, Friedrich und Goldfriedrich, Johann: Die Geschichte des deutschen Buchhandels. 4 Bde. Leipzig: Verlag des Börsenvereins der deutschen Buchhändler, 1886–1913.
- Karasch, Bernd: Critical Edition Typesetter (CET). Ein Programm zum Satz textkritischer Editionen auf PCs. In: Computergestützte Text-Edition. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu Editio 12). S. 87-99.
- Keerkhove, Derrick de: Schriftgeburten. Vom Alphabet zum Computer. München: Fink, 1995.
- Keller, Hagen u. a. (Hg.): Pragmatische Schriftlichkeit im Mittelalter. Erscheinungsformen und Entwicklungsstufen. München: Fink, 1992.
- Kett, Jürgen: MABxml-1. Ein XML Schema für das MAB2-Format.
http://www.ddb.de/standardisierung/pdf/mabxml_1_dok.pdf (18.05.2008).
- Kienecker, Michael: E-Pub – Wird das Verlagswesen durch »selfpublishing« abgelöst? In: Computergestützte Text-Edition. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu Editio 12). S. 127-133.
- Kittler, Friedrich A.: Aufschreibesysteme 1800 – 1900. München: Fink, 2003. (4. überarb. Neuaufl.).
- ders.: Protected Mode. In: Computer als Medium. Hg. von Norbert Bolz, Friedrich A. Kittler, Christoph Tholen. München: Fink, 1994. S. 209-220.
- Klook, Daniela; Angela Spahr (Hg.): Medientheorien. Eine Einführung. München: Fink, 2000.
- Koltes, Manfred: Elektronische Edition vs. Buch. Überlegungen zum Verhältnis zweier Medien zueinander am Beispiel der »Briefe an Goethe«. In: Edition und Internet. Hg. von Jörg Meier und Arne Ziegler. Berlin: Weidler, 2004. (Beiträge zur Editionsphilologie 2). S. 111-132.
- ders.: Weimarer Regestaugabe der Briefe an Goethe. In: Computergestützte Text-Edition. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu Editio 12). S. 101-115.
- Kraft, Herbert : Editionsphilologie. Mit Beitr. von Diana Schilling u. Gert Vonhoff. 2., neubearb. u. erw. Aufl. Frankfurt a. M. u. a.: Lang, 2001.
- ders. Editionsphilologie. Mit Beitr. von Jürgen Gregolin, Wilhelm Ott und Gert Vonhoff. Darmstadt: WBG, 1990.
- Kraml, Willibald und Werner-Morgenstern, Elisabeth: Computergestützte Edition einer Liederhandschrift. In: Historische Edition und Computer. Möglichkeiten und Probleme interdisziplinärer Textverarbeitung und Textbearbeitung. Hg. von Anton Schwob u. a. Graz: Leykam, 1989. S. 333-340.
- Kreutzer, Hans Joachim: Überlieferung und Edition. Textkritische und editorische Probleme, dargestellt am Beispiel einer historisch-kritischen Kleist-Ausgabe. Mit einem Beitrag von Klaus Kanzog. Heidelberg: Winter, 1976. (Beihefte zu Euphorion 7).
- Kropac, Ingo H.: Quellenbanken als Editionsmedien. In: Historische Edition und Computer. Möglichkeiten und Probleme interdisziplinärer Textverarbeitung und Textbearbeitung. Hg. von Anton Schwob u. a. Graz: Leykam, 1989. S. 243-262.
- Lemberg, Ingrid u. a. (Hg.): Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. Tübingen: Niemeyer, 2001. (Lexicographica: Series maior 107).
- Lobin, Henning (Hg.): Text im digitalen Medium. Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering. Wiesbaden: Westdeutscher Verlag, 1999.

- ders.: Textauszeichnung und Dokumentgrammatiken. In: *Texttechnologie. Perspektiven und Anwendungen*. Hg. von Henning Lobin und Lothar Lemnitzer. Tübingen: Stauffenburg, 2003. S. 51-83.
- ders. und Lemnitzer, Lothar (Hg.): *Texttechnologie. Perspektiven und Anwendungen*. Tübingen: Stauffenburg, 2003.
- Ludwig, Otto: *Geschichte des Schreibens*. Bd. 1: Von der Antike bis zum Buchdruck. Berlin: de Gruyter, 2005.
- Martens, Gunter und Woesler, Winfried (Hg.): *Edition als Wissenschaft*. Festschrift für Hans Zeller. Tübingen: Niemeyer, 1991. (Beihefte zu *Editio* 2).
- ders. und Zeller, Hans (Hg.): *Texte und Varianten. Probleme ihrer Edition und Interpretation*. München: Beck, 1971.
- ders.: *Der wohlfeile Goethe. Überlegungen zur textphilologischen Grundlegung von Leseausgaben*. In: *Edition als Wissenschaft*. Festschrift für Hans Zeller. Hg. von Gunter Martens und Winfried Woesler. Tübingen: Niemeyer, 1991. (Beihefte zu *Editio* 2). S. 72-91.
- McLuhan, Marshall: *Die Gutenberg-Galaxis. Das Ende des Buchzeitalters*. Düsseldorf/Wien: Econ, 1968.
- ders.: *The Gutenberg Galaxy: The Making of Typographic Man*. London: Routledge & Paul, 1962.
- Mehler, Alexander: *Textmining*. In: *Texttechnologie. Perspektiven und Anwendungen*. Hg. von Lothar Lemnitzer und Henning Lobin. Stauffenburg, Tübingen, 2004, S. 329-352
- ders. und Lobin, Henning (Hg.): *Automatische Textanalyse. Systeme und Methoden zur Annotation und Analyse natürlichsprachiger Texte*. Wiesbaden: VS Verlag, 2004. S. 265-283.
- dies. (Hg.): *Werkzeuge zur automatischen Analyse und Verarbeitung von Texten: Formate, Tools, Software-Systeme*. Opladen: Westdt. Verl., 2003.
- Meier, Jörg und Ziegler, Arne (Hg.): *Edition und Internet*. Berlin: Weidler, 2004. (Beiträge zur *Editionsphilologie* 2).
- ders.: *www.buch.ade? Schrift- und Lesekultur in Zeiten des Internet*. In: *Edition und Internet*. Hg. von Jörg Meier und Arne Ziegler. Berlin: Weidler, 2004. (Beiträge zur *Editionsphilologie* 2). S. 11-40.
- Meyer, Heinrich: *Edition- und Ausgabentypologie. Eine Untersuchung der editionswissenschaftlichen Literatur des 20. Jahrhunderts*. Bern, u. a.: Lang, 1992.
- Meyer-Krentler, Eckhard: *Edition & EDV. Elektronische Arbeitshilfen für Editoren, Philologen, Bücherschreiber mit dem WORD-Zusatzpaket ECCE*. München: 1992.
- Mittler, Elmar: *Zum Stand der Urheberrechtsdiskussion*. In: *Computergestützte Text-Edition*. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu *Editio* 12). S. 21-28.
- Morgenthaler, Walter: *Der produktionsorientierte Stellenkommentar in der Computer-Edition*. In: *Kommentierungsverfahren und Kommentarformen*. Hamburger Kolloquium der Arbeitsgemeinschaft für germanistische Edition 4. bis 7. März 1992, autor- und problembezogene Referate. Hg. von Gunter Martens. Tübingen: Neimeyer, 1993. (Beihefte zu *editio* 5). S. 251-255.
- ders.: *Gottfried Keller – elektronisch ediert*. In: *Jahrbuch für Computerphilologie* 1 (1999). S. 91-100.
- Morrison, Alan u. a.: *Creating and documenting electronic texts: a guide to good practice*. London: OTA, 2000.
- Müller, Jan-Dirk: *Der Körper des Buches. Zum Medienwechsel zwischen Handschrift und Druck*. In: *Materialität der Kommunikation*. Hg. von Hans U. Gumbrecht und Pfeiffer, K. Ludwig. Frankfurt a. M.: Suhrkamp, 1995. S. 203-217.
- Müller, Ulrich: *Personal Computer, Wissenschaftliche Manuskripte und Editionen*. In: *editio. Internationales Jahrbuch für Editions-wissenschaft* 2 (1988). S. 48-72.
- ders.: *Zwei Vorschläge für elektronische Bild-, Daten- und Text-Sammlungen zur mittelhochdeutschen Literatur*. In: *Historische Edition und Computer. Möglichkeiten und Probleme interdisziplinärer Textverarbeitung und Textbearbeitung*. Hg. von Anton Schwob u. a. Graz: Leykam, 1989. S. 143-149.
- Nelson, Theodor Holm: *Literary machines. The report on, and of, project Xanadu concerning word processing, electronic publishing, hypertext, thinkertoys, tomorrow's intellectual revolution, and certain other topics including knowledge, education and freedom*. Sausalito, CA: Mindful Press, 1982.

- Neuhaus, Manfred: Vom Zettelkasten zum Mikrochip. Elektronische Datenverarbeitung in der wissenschaftlichen Edition. In: Börsenblatt für den deutschen Buchhandel 152 (1985). Heft 51, S. 963-968.
- Nutt-Kofoth, Rüdiger (Hg.): Text und Edition. Positionen und Perspektiven. Berlin: Schmidt, 2000.
- ders.: Text lesen – Text sehen: Edition und Typographie. In: DVjs 78 (2004). S. 3-19.
- Oellers, Norbert und Steinecke, Hartmut (Hg.): Editionsprobleme der Literaturwissenschaft. Zeitschrift für deutsche Philologie 105 (1986). Sonderheft.
- dies. (Hg.): Probleme neugermanistischer Edition. Zeitschrift für deutsche Philologie 101 (1982). Sonderheft.
- Ott, Tobias: Datenaufbereitung für elektronische Publikationen. In: Computergestützte Text-Edition. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu Editio 12). S. 79-85.
- Ott, Wilhelm u. a.: EDV-Fibel für Editoren. Stuttgart/Tübingen: Frommann-Holzboog/Neimeyer, 1982.
- ders.: Bibliographie. EDV im Editions Wesen. In: Sprache und Datenverarbeitung 4 (1980). S. 179-184.
- ders.: Computer-unterstützte Edition. In: editio. Internationales Jahrbuch für Editions wissenschaft 3 (1989). S. 157-176.
- ders.: Der Computer als wissenschaftliches Arbeitsmittel für Editionen. In: Philosophische Editionen. Erwartungen an sie – Wirkungen durch sie. Beiträge zur VI. Internationalen Fachtagung der Arbeitsgemeinschaft philosophischer Editionen (11.-13. Juni 1992 Berlin). Hg. von Hans Gerhard Senger. Tübingen: Niemeyer, 1994 (Beihefte zu editio 6). S. 93-103.
- ders.: Edition und Datenverarbeitung. In: Herbert Kraft: Editionsphilologie. Mit Beiträgen von Jürgen Gregolin, Wilhelm Ott und Gert Vonhoff. Darmstadt: WBG, 1990. S. 59-70.
- ders.: Elektronisches Edieren. In: Geschichte der Editionsverfahren vom Altertum bis zur Gegenwart im Überblick. Ringvorlesung. Hg. von Hans-Gert Roloff. Berlin: Weidler, 2003. S. 329-357.
- Otto, Arnold: Apparateinträge und Datenbanksätze in einer Edition spätmittelalterlicher Erbauungsgedichte. Vor- und Nachteile der Automatisierung. In: Edieren in der elektronischen Ära. Hg. von Gottfried Reeg, und Martin J. Schubert. Berlin: Weidler, 2004. S. 59-69.
- Pantus, Willem-Jan: Jugendstil in Wort und Bild. Illustrierte Dichtkunst um 1900. Köln: LETTER Stiftung, 2000.
- Pflüger, Jörg: Über die Verschiedenheit des maschinellen Sprachbaus. In: Computer als Medium. Hg. von Norbert, Bolz, Friedrich A. Kittler, Christoph Tholen. München: Fink, 1994. S. 161-181.
- Plachta, Bodo und van Vliet, H. T. M.: Überlieferung, Philologie und Repräsentation. Zum Verhältnis von Editionen und Institutionen. In: Text und Edition. Positionen und Perspektiven. Hg. von Rüdiger Nutt-Kofoth. Berlin: Schmidt, 2000. S. 11-35.
- ders.: Editions wissenschaft. Eine Einführung in Methode und Praxis der Edition neuerer Texte. Stuttgart: Reclam, 1997.
- Platon: Sämtliche Werke. Berlin: Lambert Schneider, [1940].
- Raible, Wolfgang: Die Semiotik der Textgestalt. Erscheinungsformen und Folgen eines kulturellen Evolutionsprozesses. Heidelberg: Winter, 1991.
- ders. Medien-Kulturgeschichte : Mediatisierung als Grundlage unserer kulturellen Entwicklung. Vorgetragen am 11.2.2001. Heidelberg: Winter, 2006.
- ders.: Nominale Spezifikatoren (»Artikel«) in der Tradition lateinischer Juristen oder Vom Nutzen der ganzheitlichen Textbetrachtung für die Sprachgeschichte. In: Romanistisches Jahrbuch 36 (1985). S. 44-67.
- Reeg, Gottfried und Schubert, Martin J. (Hg.): Edieren in der elektronischen Ära. Berlin: Weidler, 2004.
- Reger, Ursula: Die Praxis der computerunterstützten Edition am Beispiel der Nachlaßgedichte Eichendorffs im Rahmen der historisch-kritischen Ausgabe. Möglichkeiten und Unmöglichkeiten. In: Kommentierungsverfahren und Kommentarformen. Hamburger Kolloquium der Arbeitsgemeinschaft für germanistische Edition 4. bis 7. März 1992, autor- und problembezogene Referate. Hg. von Gunter Martens. Tübingen: Niemeyer, 1993. (Beihefte zu editio 5). S. 242-250.

- Reinstein, Hagen: The DFG-Project »Retrodigitizing Johann Georg Krünitz' »Oeconomische Encyclopädie««. In: Standards und Methoden der Volltextdigitalisierung. Beiträge des internationalen Kolloquiums an der Universität Trier. 8./9. Oktober 2001. Akademie der Wissenschaften und der Literatur Mainz. Stuttgart: Steiner, 2003. S. 323-327.
- Reuß, Roland: Lesen, was gestrichen wurde. Für eine historisch-kritische Kafka-Ausgabe. In: Franz Kafka: Historisch-Kritische Ausgabe sämtlicher Handschriften und Typoskripte. Hg. von Roland Reuß. Einleitungsband. Basel/Frankfurt a. M., 1995. S. 3-25.
- ders.: Schrift & Character. In: Text. Kritische Beiträge 9 (2004). S. 185-194.
- ders.: Textkritische Editionen und Dateiformate. Notizen. In: Jahrbuch für Computerphilologie 1 (1999). S. 101-106. <http://computerphilologie.tu-darmstadt.de/jahrbuch/jb1-content.html> (18.05.2008) und <http://www.textkritik.de/technik/formate.htm> (18.05.2008).
- ders.: Zur kritischen Edition von »Der Prozess«. In: Franz Kafka: Historisch-Kritische Ausgabe sämtlicher Handschriften und Typoskripte. Hg. von Roland Reuß. Bd. 1: Der Prozess, Buch: Franz Kafka-Hefte 1. Basel/Frankfurt a. M.: Stroemfeld, 1997. S. 9-21
- Richter, Tobias: Eine TEI/SGML-Edition. (Über eine Ulysses-Edition). Dipl.Arbeit TU München 1997.
- Rischer, Thomas: Eine säurefreie elektronische Edition des Ulysses. Bestandaufnahme und einige Vorschläge. In: Schrift – Text – Edition. Hans Walter Gabler zum 65. Geburtstag. Hg. von Christiane Henkes. Tübingen: Niemeyer, 2003. (Beihefte zu Editio 19). S. 339-349.
- Robinson, Peter und Taylor, Kevin: Publishing an Electronic Textual Edition: The Case of The Wife of Bath's Pologue on CD-ROM. In: Computers and the humanities. Official journal of the Association for Computers and the Humanities. Special issue on Electronic publishing 32 (1998). S. 271-284.
- Roloff, Hans-Gert (Hg.): Geschichte der Editionsverfahren vom Altertum bis zur Gegenwart im Überblick. Ringvorlesung. Berlin: Weidler, 2003.
- Rösler, Uta: Ein mittelhochdeutsches Wörterbuch auf CD-ROM – Strukturbeschreibung der Wörterbuchartikel in Matthias Lexers »Mittelhochdeutschem Handwörterbuch« für die Herstellung einer elektronischen Version auf CD-ROM. Schriftliche Hausarbeit zur Akademischen Abschlußprüfung (Magisterarbeit) im Fach Germanistik. Universität Trier: Juni 1998.
- Sahle, Patrick: »Sinnsuche in der Badewanne«. Tagungsbericht: Standards und Methoden der Volltextdigitalisierung: Trier, 8. und 9. Oktober 2001. In: Jahrbuch für Computerphilologie 4 (2002). S. 191-195.
- ders.: Digitale Editionstechniken und historische Quellen. In: Internet-Handbuch Geschichte. Hg. von Stuart Jenks und Stephanie Marra. Köln u. a.: Böhlau, 2001. S. 153-166.
- ders.: Digitales Archiv und Digitale Edition. Anmerkungen zur Begriffsklärung. In: Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Hg. von Michael Stolz u. a. Basel: Germanistik.ch, 2007. http://www.germanistik.ch/scripts/download.php?id=Digitales_Archiv_und_digitale_Edition (18.05.2008).
- ders.: Vom editorischen Fachwissen zur digitalen Edition: Der Editionsprozeß zwischen Quellenbeschreibung und Benutzeroberfläche. In: Fundus – Forum für Geschichte und ihre Quellen Heft 2.
- Scheibe, Siegfried: Kleine Schriften zur Editionswissenschaft. Berlin: Weidler, 1997. (Berliner Beiträge zur Editionswissenschaft.)
- ders.: Zu einigen Grundprinzipien einer historisch-kritischen Ausgabe. In: Texte und Varianten. Probleme ihrer Edition und Interpretation. Hg. von Gunter Martens und Hans Zeller. München: Beck, 1971. S. 1-44.
- ders.: Zur Darstellung der Überlieferung in historisch-kritischen Editionen. In: Edition als Wissenschaft. Festschrift für Hans Zeller. Hg. von Gunter Martens und Winfried Woesler. Tübingen: Niemeyer, 1991. (Beihefte zu Editio 2). S. 17-30.
- Schmidt, Ernst-Joachim (Hg.): Kritische Bewahrung. Beiträge zur deutschen Philologie. Festschrift für Werner Schröder zum 60. Geburtstag. Berlin: Erich Schmidt, 1974.

- Schmidt, Frieder: Neuland für die Buchgeschichte – Quellenaufbereitung im Zeitalter des WWW. Hypertext Markup Language (HTML), Standard Generalized Markup Language (SGML) und die Guidelines for Electronic Text Encoding and Interchange der Text Encoding Initiative (TEI). In: Leipziger Jahrbuch zur Buchgeschichte 7 (1997). S. 343-365.
- Schmidt, Ingrid und Müller, Carolin: Entwicklung eines lexikographischen Modells: Ein neuer Ansatz. In: Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. Hg. von Ingrid Lemberg u. a. Tübingen: Niemeyer, 2001. (Lexicographica: Series maior 107). S. 29-52.
- Schmidt, Ingrid: Modellierung von Metadaten. In: Texttechnologie. Perspektiven und Anwendungen. Hg. von Henning Lobin und Lothar Lemnitzer. Tübingen: Stauffenberg, 2004. S. 143-164.
- Schmitt, Alfred: Die Erfindung der Schrift. Rede, geh. vor d. Univ. Erlangen am 19. Juni 1938 bei d. Eröffnungsfeier d. Universitätswoche. Erlangen: Palm & Enke, 1938. (Erlanger Universitäts-Reden 22.)
- Schnarr, Hermann: Umstellung einer Edition auf EDV – am Beispiel der Edition des Nikolaus von Kues. In: Philosophische Editionen. Erwartungen an sie – Wirkungen durch sie. Beiträge zur VI. Internationalen Fachtagung der Arbeitsgemeinschaft philosophischer Editionen (11.-13. Juni 1992 Berlin). Hg. von Hans Gerhard Senger. Tübingen: Niemeyer, 1994 (Beihefte zu editio 6). S. 104-110.
- Schneider, Ute: Neuprofilierung eines Autors: Karl F. Gutzkows Schriften und Briefe im Internet. In: Edition und Internet. Hg. von Jörg Meier und Arne Ziegler. Berlin: Weidler, 2004. (Beiträge zur Editionsphilologie 2). S. 133-148.
- Scholler, Harald: A word index to the Nibelungenklage. Based on the K. Lachmann ed. Ann Arbor: Univ. of Michigan Pr., 1966.
ders.: Über die Förderung der Nibelungenforschung durch Elektronenrechner. In: Zeitschrift für deutsches Altertum 95 (1966). S. 149-165.
- Schulz, Klaus U. und Meuss, Holger: Abfrage strukturierter Dokumente. In: Texttechnologie. Perspektiven und Anwendungen. Hg. von Henning Lobin und Lothar Lemnitzer. Tübingen: Stauffenberg, 2003. S. 379-406.
- Schwob, Anton u. a. (Hg.): Historische Edition und Computer. Möglichkeiten und Probleme interdisziplinärer Textverarbeitung und Textbearbeitung. Graz: Leykam, 1989.
- Seelbach, Ulrich: Avancinis »Pietas victrix«. Werkstattbericht zur Erstellung einer computerunterstützten Edition. In: Probleme der Edition von Texten der Frühen Neuzeit. Beiträge zur Arbeitstagung der Kommission für die Edition von Texten der Frühen Neuzeit. Hg. von Lothar Mundt, Hans-Gert Roloff und Ulrich Seelbach. Tübingen: Niemeyer, 1992 (Beihefte zu editio 3). S. 168-177.
- Seib, Renate: Exilpresse Digital. Deutschsprachige Exilzeitschriften 1933–1945. In: Fundus – Forum für Geschichte und ihre Quellen Heft 5. <http://webdoc.gwdg.de/edoc/p/fundus/5/seib.pdf> (18.05.2008).
- Shillingsburg, Peter: Anglo-Amerikanische Editionswissenschaft. Ein knapper Überblick. In: Text und Edition. Positionen und Perspektiven. Hg. von Rüdiger Nutt-Kofoth. Berlin: Schmidt, 2000. S. 143-164.
- Stängle, Peter: Gewinne und Verluste. In: Text. Kritische Beiträge 9 (2004). S. 181-184.
- Steer, Georg: Edieren mit Hilfe der EDV. Zur Textausgabe der »Rechtssumme« Bruder Bertholds. In: Maschinelle Verarbeitung altdeutscher Texte. Beiträge zum Symposium 3 (1977 [1980]). S. 121-130.
- Stein, Peter: Schriftkultur. Eine Geschichte des Schreibens und Lesens. Darmstadt: WBG, 2006.
- Stolz, Michael u. a. (Hg.): Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Bern: germanistik.ch, 2005.
http://www.germanistik.ch/buch.php?id=Literatur_Neue_Medien (18.05.2008). Druckfassung: Basel: germanistik.ch, 2007.
ders.: Computergestütztes Kollationieren. Ein Werkstattbericht aus dem Basler Parzival-Projekt. In: Edieren in der elektronischen Ära. Hg. von Gottfried Reeg, und Martin J. Schubert. Berlin: Weidler, 2004. S. 113-126.
- Storrer, Angelika: Text und Hypertext. In: Texttechnologie. Perspektiven und Anwendungen. Hg. von Henning Lobin und Lothar Lemnitzer. Tübingen: Stauffenberg, 2003. S. 13-50.
- Stüben, Jens: Edition und Interpretation. In: Text und Edition. Positionen und Perspektiven. Hg. von Rüdiger Nutt-Kofoth. Berlin: Schmidt, 2000. S. 263-302.

- Thaller, Manfred (Hg.): Digitale Bausteine für die geisteswissenschaftliche Forschung. Göttingen: Duerkohn & Radicke, 2003. (Fundus Beiheft 5).
- ders.: Datenbanken als Editionsformen? In: Historische Edition und Computer. Möglichkeiten und Probleme interdisziplinärer Textverarbeitung und Textbearbeitung. Hg. von Anton Schwob u. a. Graz: Leykam, 1989. S. 215-242.
- ders.: Secundum Manus. Zur Datenverarbeitung mehrschichtiger Editionen. In: Geschichte und ihre Quellen. Festschrift für Friedrich Hausmann zum 70. Geb. Hg. von Reinhard Härtel. Graz: Akad. Dr. u. Verl. Anst., 1987. S. 629-638.
- Tiling, Nikolaus: Einsatz von EDV bei der Nachlaß-Erschließung am Beispiel Hubert Fichte. In: Kommentierungsverfahren und Kommentarformen. Hamburger Kolloquium der Arbeitsgemeinschaft für germanistische Edition 4. bis 7. März 1992, autor- und problembezogene Referate. Hg. von Gunter Martens. Tübingen: 1993 (Beihefte zu editio 5). S. 256-261.
- Tollenare, Felician de: Lexikographie mit Hilfe des elektronischen Informationswandlers. In: Zeitschrift für deutsche Sprache 21 (1965), Heft 1/1, S. 1-19.
- Urchueguía, Christina: Edition und Faksimile. Versuch über die Subjektivität des Objektivs. In: Text und Edition. Positionen und Perspektiven. Hg. von Rüdiger Nutt-Kofoth. Berlin: Schmidt, 2000. S. 323-352.
- Waldenberger, Sonja: Chancen und Möglichkeiten der Hypermedialisierung. In: Edition und Internet. Hg. von Jörg Meier und Arne Ziegler. Berlin: Weidler, 2004. (Beiträge zur Editionsphilologie 2). S. 149-166.
- Wallmannsberger, Josef: Virtuelle Textwelten. Theoretische Grundlagen und Implementationsperspektiven der anglistischen Computerphilologie. Heidelberg: Winter, 1994.
- Weber, Hartmut: Archiv-Server/Server Archive – Wie sehen die Kulturspeicher der Zukunft aus? In: Computergestützte Text-Edition. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu Editio 12). S. 135-141.
- Wedeles, Lauren: Prof. Nelson Talk. Analyzes P.R.I.D.E. In: Vassar Miscellany News, 3.2.1965. S. 3 f.
- Wegstein, Werner: Zeichenerkennung und elektronische Textfassung. In: Philosophische Editionen. Erwartungen an sie – Wirkungen durch sie. Beiträge zur VI. Internationalen Fachtagung der Arbeitsgemeinschaft philosophischer Editionen (11.-13. Juni 1992 Berlin). Hg. von Hans Gerhard Senger. Tübingen: Niemeyer, 1994 (Beihefte zu editio 6). S. 118-124.
- Wehde, Susanne: Typographische Kultur. Eine zeichentheoretische und kulturgeschichtliche Studie zur Typographie und ihrer Entwicklung. Tübingen: Niemeyer, 2000. (Studien und Texte zur Sozialgeschichte der Literatur 69).
- Wender, Herbert und Peter, Robert: Probleme der Wiederverwendung elektronisch gespeicherter Texte. Zwei Fallstudien. In: Computergestützte Text-Edition. Hg. von Roland Kamzelak. Tübingen: Niemeyer, 1999. (Beihefte zu Editio 12). S. 47-60.
- Wirtz, Thomas: Stirb viermal, jetzt erst recht. Die Kompanie »Marburg« der Büchner-Forschung zog ins Feld und kehrt mit vollen Schulranzen zurück. Frankfurter Allgemeine Zeitung Nr. 289, 12.12.2000.
- Wisbey, Roy: Computer und Philologie in Vergangenheit, Gegenwart und Zukunft. In: Maschinelle Verarbeitung altdeutscher Texte. Beiträge zum Symposium 4 (1988 [1991]). S. 346-361.
- ders.: Die Bedeutung des Computers für die Edition altdeutscher Texte, in: Kritische Bewahrung. Beiträge zur deutschen Philologie. Festschrift für Werner Schröder zum 60. Geburtstag. Hg. von Ernst-Joachim Schmidt. Berlin: Erich Schmidt, 1974. S. 344-356.
- ders.: Ein computerlesbares Textarchiv des Frühmittelhochdeutschen. In: Jahrbuch für Internationale Germanistik 7 (1969). Heft 2, S. 37-46.
- ders.: Vollständige Verskonkordanz zur Wiener Genesis. Mit einem rückläufigen Wörterbuch zum Formenbestand. Berlin: E. Schmidt, 1967.
- Wittmann, Reinhard: Geschichte des deutschen Buchhandels. München: Beck, 1999.
- Yan, Zhenjiang: Schriftsystem, Literalisierung, Literalität. Frankfurt a. M. u. a.: Lang, 2000.

-
- Zeller, Hans :Befund und Deutung. Interpretation und Dokumentation als Ziel und Methode der Edition. In: Texte und Varianten. Probleme ihrer Edition und Interpretation. Hg. von Gunter Martens und Hans Zeller. München: Beck, 1971. S. 45-89.
- ders.: Textwahl für die Datenverarbeitung. Der Wortindex zu Werken Gottfried Kellers und die Keller-Ausgabe. In: Euphorion 66 (1972). S. 383-396.
- Zwerschina, Hermann: Variantenverzeichnung, Arbeitsweise des Autors und Darstellung der Textgenese. In: Text und Edition. Positionen und Perspektiven. Hg. von Rüdiger Nutt-Kofoth. Berlin: Schmidt, 2000. S. 203-229.

Internetressourcen

Portale und Homepages

Arbeitsstelle Kant-Ausgabe der Göttinger Akademie der Wissenschaften

http://web.uni-marburg.de/kant/webseitn/home_inh.htm (18.05.2008).

Bayerische Staatsbibliothek, Digitale Sammlungen

http://mdz.bib-bvb.de/digbib/lexika/@Generic__CollectionView;cs=default;ts=default;lang=de (18.05.2008).

Carsten Worm Infosoftware

<http://www.infosoftware.de/> (18.05.2008).

Deutsche Nationalbibliothek

<http://www.d-nb.de/> (18.05.2008).

Digital Library Forum

<http://www.dl-forum.de> (18.05.2008).

Directmedia Publishing

<http://www.digitale-bibliothek.de/> (18.05.2008).

Docbook

<http://www.docbook.org> (18.05.2008).

Dublin Core Methadata initiative

<http://www.dublincore.org> (18.05.2008).

Google

<http://www.google.de> (18.05.2008).

Google Books

<http://www.google.de/books> (18.05.2008).

Heinrich Heine Portal

<http://germazope.uni-trier.de/Projects/HHP/> (18.05.2008).

Die Historisch-Kritische Gottfried Keller-Ausgabe (HKKA)

<http://www.gottfriedkeller.ch/hkka/hkka.htm> (18.05.2008).

International Digital Publishing Forum

<http://www.idpf.org/> (18.05.2008).

International Tustep User Group (ITUG)

<http://www.itug.de/> (18.05.2008).

Organization for the Advancement of Structured Information Standards (OASIS)

<http://www.oasis-open.org/home/index.php> (18.05.2008).

O'Reilly xml.com

<http://www.xml.com> (18.05.2008).

Project Xanadou

<http://xanadu.com> (18.05.2008).

Projekt Gutenberg

<http://gutenberg.spiegel.de> (18.05.2008).

Text Encoding Initiative (TEI)

<http://www.tei-c.org> (18.05.2008).

Tustep Homepage

<http://www.zdv.uni-tuebingen.de/tustep/index.html> (18.05.2008).

W₃C Cascading Style Sheets home page

<http://www.w3.org/Style/CSS/> (18.05.2008).

W₃C The Extensible Stylesheet Language Family (XSL)

<http://www.w3.org/Style/XSL/> (18.05.2008).

- W3C XML Path Language (XPath) 2.0
<http://www.w3.org/TR/xpath20/> (18.05.2008).
- W3C XML Path Language (XPath) Version 1.0
<http://www.w3.org/TR/xpath/> (18.05.2008).
- W3C XML Pointer, XML Base and XML Linking
<http://www.w3.org/XML/Linking/> (18.05.2008).
- W3C XML Query (XQuery)
<http://www.w3.org/XML/Query/> (18.05.2008).
- World Wide Web Consortium (W3C)
<http://www.w3.org/> (18.05.2008).
- Word Cruncher
<http://wordcruncher.byu.edu/wordcruncher/default.htm> (18.05.2008).
- Zeno.org
<http://www.zeno.org/> (18.05.2008).

Texte und Textausgaben im W3

- A Gentle Introduction to XML.
<http://www.tei-c.org/release/doc/tei-p4-doc/html/SG.html> (18.05.2008).
- Adelung, Grammatisch-kritisches Wörterbuch bei der bayerischen Staatsbibliothek.
<http://mdz.bib-bvb.de/digbib/lexika/adelung> (18.05.2008).
- Altrichter, Helmut: Retrodigitalisierung in Deutschland – Versuch einer Zwischenbilanz.
<http://www.bsb-muenchen.de/mdz/forum/altrichter/> (18.05.2008).
- Bradley, John und Rockwell, Geoffrey: TACT Web 1.0 Workbook.
<http://tactweb.humanities.mcmaster.ca/tactweb/doc/twintro.htm> (18.05.2008).
- Brechts Werke im WWW.
<http://brecht.chadwyck.com> (18.05.2008).
- Burnard, Lou und Bauman, Syd: TEI P5: Guidelines for Electronic Text Encoding and Interchange. TEI Consortium 2007.
<http://www.tei-c.org/Guidelines/P5/> (18.05.2008).
- Das Deutsche Wörterbuch von Jacob und Wilhelm Grimm auf CD-ROM und im Internet.
<http://germazope.uni-trier.de/Projects/DWB> (18.05.2008).
- Eversberg, Bernhard: Was sind und was sollen Bibliothekarische Datenformate. Überarbeitete und erweiterte Neuauflage. WWW-Version mit Ergänzungen.
<http://www.allegro-c.de/formate/formate.htm> (18.05.2008).
- Exilpresse Digital. Deutsche Exilzeitschriften 1933–1945
<http://deposit.ddb.de/online/exil/exil.htm> (18.05.2008).
- Fundus – Forum für Geschichte und ihre Quellen. Hg. von Manfred Thaller.
<http://webdoc.gwdg.de/edoc/p/fundus/index.html> (18.05.2008).
- Gabler, Hans-Walter: Rezension: Roland Kamzelak (Hg.): Computergestützte Text-Edition.
<http://computerphilologie.uni-muenchen.de/jgo1/gabler.html> (18.05.2008).
- Goethe, Johann Wolfgang von: Goethes Werke im WWW.
<http://goethe.chadwyck.com/> (18.05.2008).
- Gottfried Keller (Historisch-Kritische Gottfried Keller Ausgabe).
<http://www.gottfriedkeller.ch/> (18.05.2008).
- Hildenbrandt, Vera und Schares, Thomas: Das Grimmsche Wörterbuch geht ins 21. Jahrhundert: Präsentation eines Prototyps des digitalen Deutschen Wörterbuchs von Jacob und Wilhelm Grimm.
<http://www.bsb-muenchen.de/mdz/forum/hildenbrandt/index.htm> (18.05.2008).
- Index Thomisticus.
<http://www.corpusthomicum.org> (18.05.2008).

- Jahrbuch für Computerphilologie.
<http://www.computerphilologie.de/> (18.05.2008).
- Jannidis, Fotis: TEI in der Praxis. Arbeitsbericht.
<http://computerphilologie.tu-darmstadt.de/praxis/teipraxis.html> (18.05.2008).
- Kafka, Franz: Kafkas Werke im WWW.
<http://kafka.chadwyck.com> (18.05.2008).
- Kett, Jürgen: MABxml-1. Ein XML Schema für das MAB2-Format.
http://www.ddb.de/standardisierung/pdf/mabxml_1_dok.pdf (18.05.2008).
- Kritisches Lexikon zur deutschsprachigen Gegenwartsliteratur.
<http://www.klgonline.de>. (18.05.2008).
- Das Microsoft Office Open XML-Format.
<http://www.microsoft.com/germany/msdn/office/openxml/default.mspx> (18.05.2008).
- Open Document Format for Office Applications (OpenDocument) v1.1.
<http://docs.oasis-open.org/office/v1.1/OS/OpenDocument-v1.1-html/OpenDocument-v1.1.html>
(18.05.2008)
- Mittelhochdeutsche Wörterbücher im Verbund
<http://gaer27.uni-trier.de/MWV-online/MWV-online.html> (18.05.2008).
- Reuß, Roland: Textkritische Editionen und Dateiformate. Notizen. In: Jahrbuch für Computerphilologie 1 (1999). S. 101-106.
<http://computerphilologie.tu-darmstadt.de/jahrbuch/jb1-content.html> (18.05.2008) und
<http://www.textkritik.de/technik/formate.htm> (18.05.2008).
- Rezensionsorgane und Literaturzeitschriften: Retrospektive Digitalisierung wissenschaftlicher
Rezensionsorgane und Literaturzeitschriften des 18. und 19. Jahrhunderts aus dem deutschen Sprachraum.
<http://www.ub.uni-bielefeld.de/diglib/aufklaerung/index.htm> (18.05.2008).
- Sahle, Patrick: Digitales Archiv und Digitale Edition. Anmerkungen zur Begriffsklärung.
http://www.germanistik.ch/scripts/download.php?id=Digitales_Archiv_und_digitale_Edition (18.05.2008).
- Schillers Werke im WWW.
<http://schiller.chadwyck.com> (18.05.2008).
- Sperberg-McQueen, C. M. und Burnard, Lou: TEI P4. Guidelines for Electronic Text Encoding and
Interchange. XML-compatible edition. TEI Consortium 2001–2004.
<http://www.tei-c.org/P4X/> (18.05.2008).
- Zedler, Johann Heinrich: Grosses vollständiges Universal-Lexicon aller Wissenschaften und Künste. Leipzig
1732-1754.
<http://mdz10.bib-bvb.de/~zedler/zedler2007/index.html> (18.05.2008).