

DIE VORSICHTIGE ANNÄHERUNG EINES INFORMATIKERS AN DIE EDITIONSWISSENSCHAFT

KLAUS MEYER-WEGENER

Als ich gegen Ende des Jahres 2012 gebeten wurde, im Interdisziplinären Zentrum für Editionswissenschaften (IZED) der FAU Erlangen-Nürnberg mitzuwirken, hatte ich im Grunde genommen keine Vorstellung davon, was Editionswissenschaft überhaupt ist. Es interessiert mich aber durchaus, weil ich einerseits ein großer Bücherfreund bin und mich andererseits aber auch beruflich schon einige Zeit mit Multimedia-Datenbanken¹ und Digitalen Bibliotheken² befasst habe. Das IZED legte von Anfang an großen Wert auf die Einbeziehung der Informatik. Editionen können sich heute einfach nicht mehr auf das gedruckte Buch beschränken. Man denkt sofort an einen Datenträger (CD, DVD) oder einen Web-Auftritt. Damit ist aber überhaupt noch nicht festgelegt, wie eine Edition dann auszu-sehen hätte und wie der Umgang damit erfolgen soll.

An mich wird oft die Frage gestellt, was die Informatik hier denn nun an Unterstützung bieten kann. Es ist dabei sicher nicht allein eine Frage der Technik, sondern auch der Gestaltung, und da sind Informatiker notorisch überfordert. Deshalb werde ich mich dazu auch besser nicht äußern. Hier müssen Konventionen erst noch entwickelt werden, etwa zu Fragen wie: Was sieht man gleichzeitig und was erst bei Bedarf? Wie und wo werden die verschiedenen Bestandteile dargestellt?

In diesem Beitrag wird es also darum gehen, was die Informatik der Editions-wissenschaft anbieten kann. Damit werden keine Editionen von selbst (automatisch) entstehen! Der Aufwand wird vermutlich am Anfang sogar noch größer sein als zuvor, weil man sich als Herausgeber erst in die Technik einarbeiten muss.

Das lohnt sich deshalb, weil neue Möglichkeiten entstehen: Man kann verschiedene Ansichten für unterschiedliche Benutzerkreise erzeugen, die sich auf denselben Datenbestand beziehen. Eine solche Wiederverwendbarkeit von Daten war immer eine treibende Kraft bei der Entwicklung und beim Einsatz von Datenbanken und ist deshalb

1 Klaus Meyer-Wegener: Multimediale Datenbanken : Einsatz von Datenbanktechnik in Multimedia-Systemen, Stuttgart 2003.

2 Klaus Meyer-Wegener: Alexandria 2.0 – Zur digitalen Zukunft der Bibliotheken, in: Klimawandel. Mediengesellschaft, hg. von Karl Möseneder, Erlangen 2012 (Erlanger Forschungen, Reihe A 124/125), S. 129-146.

recht gut untersucht worden. Des Weiteren kommen als Illustrationen nicht nur Bilder in Frage, sondern auch Ton und Video. Man könnte sogar Daten in Form von Excel-Tabellen hinzufügen, die dem Leser Nachvollziehbarkeit bieten und die Durchführung eigener Auswertungen erlauben. Man denke nur an Rechnungsbücher oder andere Wirtschaftszugnisse.

Begonnen hat die Erstellung digitaler Editionen meist mit Textverarbeitung und Desktop Publishing, doch sind diese Systeme viel zu stark auf Layout ausgerichtet und zu wenig auf Inhalt und Struktur, was die Wiederverwendung stark einschränkt. Das gilt auch für die Hypertext Markup Language (HTML), die als die Textbeschreibungssprache des World-wide Web (WWW) inzwischen als allgemein bekannt vorausgesetzt werden kann³. Sie ist eigentlich nur ein Spezialfall von Hypertext, also der Möglichkeit, in elektronisch gespeicherten Texten Verweise (Links) auf andere Texte einzustreuen, denen man mit Systemunterstützung sehr einfach folgen kann. Mit der Verbreitung von HTML ist die Vielfalt der früheren Hypertext-Systeme⁴ praktisch untergegangen. Die Möglichkeiten der Strukturierung sind nun nur noch relativ elementar. Das Prinzip der elektronischen Querverweise hat sich damit aber durchgesetzt.

Je mehr Web-Seiten und Web-Auftritte mit HTML erstellt wurden, desto drängender wurde der Wunsch nach Unterstützung von Wiederverwendung (Re-use), damit wiederkehrende Bausteine nicht immer wieder neu erstellt werden müssen. Was einmal erfasst wurde, sollte nicht (in einem anderen Format) noch einmal erfasst werden und auch nicht von Hand umformatiert werden müssen. Das gelingt leider in der Realität noch viel zu selten. Und es macht am Anfang auch ein wenig mehr Arbeit, weil man mehr sagen bzw. eingeben muss und den Nutzen unmittelbar noch gar nicht sieht.

Damit sind wir bei Datenbanken bzw. Content-Management-Systemen (CMS)⁵, die genau hier Unterstützung bieten. Sie speichern Bausteine für Texte und Web-Seiten als ›Inhalte‹, daher der Name. Ziel ist hier, dass jeder Inhalt nur einmal vorhanden sein muss, aber in viele verschiedene Web-Seiten ›eingebaut‹ werden kann. Ändert man den Inhalt, so wirkt sich das auf alle Web-Seiten aus, die ihn enthalten – was bei Logos, Adressen, Ansprechpartnern und gesetzlich vorgeschriebenen Texten sehr nützlich und zeitsparend sein kann. Auch bei Editionen kann das sehr sinnvoll sein, um ein einheitliches Erscheinungsbild zu gewährleisten.

Generell ist eine Trennung von Struktur und Inhalt einerseits und Layout oder Gestaltung andererseits sehr wünschenswert, weil das Layout in den verschiedenen Präsentationsformen (Druck, elektronisches Seitenformat wie PDF, e-Book, Hypertext) sehr unterschiedlich ausfallen kann, die Inhalte aber gleich bleiben sollen. Hier kommen wie-

3 SELFHTML, hg. von SELFHTML e.V., URL: www.selfhtml.org [zuletzt besucht am 13.11.2015].

4 Jeff Conklin: Hypertext: An Introduction and Survey, in: IEEE Computer 20/9 (1987), S. 17-41.

5 Stefan Spörrer: Content Management Systeme: Begriffsstruktur und Praxisbeispiel, Köln 2009.

der die unterschiedlichen Ansichten zum Tragen, von denen oben bereits die Rede war. Man kann die Inhalte für bestimmte Benutzergruppen (interessierte Laien, Fachleute) unterschiedlich auswählen und in geeigneter Form präsentieren.

Der Inhalt besteht aus den oben bereits erwähnten Bausteinen, also Texten und Abbildungen, die aber auch noch in einer Struktur angeordnet werden. Sie treten als Absatz auf oder als Überschrift oder als Fußnote, um nur einige Beispiele zu nennen. Zur Definition einer solchen Struktur hat sich inzwischen die eXtensible Markup Language (XML)⁶ etabliert. Speziell für die Geistes- und Sozialwissenschaften wurde die Text Encoding Initiative (TEI)⁷ ins Leben gerufen, die mit den Mitteln der XML immer wieder vorkommende Strukturen in einheitlicher Weise definiert.

Damit kann man nun Texte in einer vom Layout noch unabhängigen Form strukturieren und auch mit multimedialen Elementen anreichern, also Illustrationen, Porträts, Landkarten, Stadtplänen, bis hin zu 3D-Modellen von Gebäuden, die es heute vielleicht gar nicht mehr gibt oder nur als Ruinen. Um dann zu einer Edition zu gelangen, muss man die Texte aber geeignet darstellen, also dann eben doch ein Layout hinzufügen. Dafür gibt es Mittel wie die ›Cascading Style Sheets‹ (CSS)⁸, die zu bestimmten Elementen der Struktur Darstellungsregeln definieren. Damit sagt man dann z. B., dass eine Überschrift immer groß und fett gedruckt werden soll, dass davor ein paar Zeilen Leerraum stehen sollen oder gar ein Seitenwechsel usw. Man kann demselben Text nacheinander unterschiedliche Darstellungsregeln zuordnen und so ganz unterschiedliche Darstellungen desselben Textes erreichen. Damit steht die erforderliche Technik zur Verfügung; was aber unverändert schwierig bleibt ist die Frage, welche Gestaltung man denn bei der elektronischen Edition wählen sollte. Für den Druck auf Papier haben sich seit Jahrhunderten Konventionen entwickelt, die den Lesern vertraut sind und die ihnen einen schnellen Zugang zu den Textbestandteilen gewähren, die sie besonders interessieren. Für elektronische Präsentation gibt es diese Konventionen aber erst in Ansätzen. Man probiert vieles einfach aus und erzeugt dadurch oft eine verwirrende Vielfalt, die den Lesern nun viel Zeit zum Eingewöhnen abverlangt.

Ein Problem scheint dabei langsam zu verschwinden: die Weigerung, längere Texte am Bildschirm zu lesen. Die Bildschirme sind inzwischen sehr viel besser geworden und erreichen nahezu die Schärfe und auch den Kontrast von Druck auf Papier. Gerade

6 Helmut Vonhoegen: Einstieg in XML. Aktuelle Standards: XML Schema, XSL, XLink, Bonn ²2013.

7 TEI: www.tei-c.org [zuletzt besucht am 13.11.2015].

8 Ingo Chao, Corina Rudel: Fortgeschrittene CSS-Techniken, Bonn ³2012.

die junge Generation wächst geradezu mit E-Book-Readern auf, und die ältere freut sich auch, wenn sie die Schrift vergrößern kann – was im Druck nun mal nicht möglich ist.

Viele sind aber mit dem reinen Lesen allein gar nicht mehr zufrieden, sondern möchten auch noch Hintergrundinformation unterschiedlichen Typs abrufen können. So möchte man bei Übersetzungen an manchen Stellen doch auch mal das Original sehen oder bei historischen Quellen die Darstellung des Textbilds. Im Druck geht das nur durch Synopsen, die viel Platz in Anspruch nehmen – auch wenn man sie zeitweise gar nicht benötigt. Elektronisch können sie im Hintergrund bleiben, bis man sie explizit abrufft. Das gilt trivialerweise auch für Fußnoten, die nun den Seiteninhalt nicht mehr verkleinern. Der Text liegt gleichsam in Schichten übereinander, immer mit der Möglichkeit, auf die darunterliegenden Schichten zurückzugreifen, wenn man das möchte. Innerhalb einer Schicht kann man ansonsten aber mit ungestörter Lektüre durch den Text gehen.

Und dann kommt sehr schnell der Wunsch auf, noch mehr über den Text zu erfahren, gerade bei Wissenschaftlern als Lesern. Das führt dann zu den sog. Metadaten, die ebenfalls mit dem Text erfasst werden können – aber eben auch erfasst werden müssen, was durchaus einiges an Arbeit macht. Es gibt hier eine unglaubliche Vielfalt, von den Informationen über die Quelle (dafür hat die oben genannte TEI eine ganze Reihe von Optionen vereinheitlicht) bis hin zu weiterführenden Angaben zu Autoren oder zu den zentralen Aussagen des Textes, die Personen, Orten, Zeitpunkte oder -intervalle, Gebäude, Kunstgegenstände und noch so vieles mehr betreffen können. Gerade wenn man alte oder fremde Texte liest, weiß man ja oft gar nicht, was man damals mit einem Wort oder Ausdruck verbunden hat und wie man das heute sagen würde, um die gleiche Wirkung zu erreichen. Und es geht sogar über den vorliegenden Text hinaus, wenn man fragt: Wer hat denn noch darüber geschrieben, und wie? Hier muss man die Texte mit strukturierten Daten verknüpfen, die solche Informationen aufnehmen können und sie damit abrufbar machen.

Dafür bietet die Informatik eine ganze Reihe von Ausdrucksmöglichkeiten an, die allgemein als Wissensrepräsentation bezeichnet werden. Ein wichtiger Bestandteil sind Begriffssystematiken, die etwas unglücklich ›Ontologien‹ genannt werden⁹. Diesem Anspruch können sie wohl kaum gerecht werden, aber sie tragen sehr zu einer Vereinheitlichung von Begriffsverwendungen bei, die in der Wissenschaft immer benötigt wird und die hier nun explizit und formal erfasst werden kann. Im Kontext des oben eingeführten World-wide Web (WWW) hat man dafür später dann den Begriff ›Semantic Web‹ eingeführt¹⁰. Hier benutzt man als Wissensrepräsentation das Resource Description Framework (RDF), das elementare Aussagen der Form ›Subjekt – Prädikat – Objekt‹, sog. Tri-

9 Steffen Staab, Rudi Studer: Handbook on Ontologies, Berlin/Heidelberg, 2004.

10 Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, York Sure: Semantic Web. Grundlagen, Berlin/Heidelberg 2008.

pel, vorsieht, die miteinander verknüpft werden können, indem das Objekt der einen Aussage als Subjekt einer anderen verwendet wird. Eine große Menge derart verknüpfter Elementaraussagen erlaubt dann schon die Formulierung recht komplexer Tatbestände. Die dazu passende Ontologie heißt OWL für ›Web Ontology Language‹ – nein, da sind im Akronym die beiden ersten Buchstaben nicht fehlerhaft vertauscht, sondern ganz bewusst: ein (typischer) Informatik-Scherz.

Eine digitale Edition kann also heute aus einer sehr großen und heterogenen Menge von Daten bestehen. Man beginnt typischerweise mit einer Quelle und reichert sie dann immer mehr an. Am Anfang steht dabei oft – nach der Lektüre des Dokumentes selbst, die keinesfalls unterlassen werden sollte – das Scannen, also schlicht das Abfotografieren der Seiten, mit anschließender optischer Zeichenerkennung (Optical Character Recognition, OCR). Damit hat man aber nur das Rohmaterial und noch lange keine Edition. Das Foto des Textes kann durchaus auch Bestandteil einer Edition sein, das wurde oben ja schon erwähnt. Eine Transkription geht dann aber noch deutlich über den Rohtext hinaus, und das meint nicht nur die Korrekturen, die das Resultat einer OCR mit Sicherheit noch verlangt. Damit beginnen dann auch die Metadaten mit den Antworten auf Fragen wie: Warum steht das im Text, und warum so? Wie wurde der Text zu seiner Zeit verstanden? Wie würde man ihn heute schreiben, damit er wieder so verstanden wird? Und hier beginnt auch die Rolle des menschlichen Editors, denn solche Hinzufügungen lassen sich kaum automatisieren. Die Technik erlaubt dann wieder die geeignete Speicherung und die Nutzung durch andere.

Gerade bei digitalen Editionen will man aber eigentlich keinen Zustand erreichen, in dem die Edition dann einfach ›fertig‹ ist. Lesern, besonders Wissenschaftlern, aber auch interessierten Laien wird zu einer Edition immer noch etwas auffallen oder einfallen, was man hinzufügen oder zumindest einmal zur Diskussion stellen könnte. Das ist bei einer digitalen Edition ja auch kein Problem, wie das sog. Web 2.0 demonstriert hat, zum Beispiel mit den sog. Wikis. Hier können beliebige Benutzer Änderungen vornehmen. Allerdings hat man dabei auf die Nachvollziehbarkeit geachtet: Änderungen werden stets protokolliert und sind ggf. auch sehr einfach rückgängig zu machen. Zu jeder neuen Aussage wird festgehalten, von wem sie hinzugefügt worden ist. Das letzte Wort hat dann immer noch der Herausgeber. In vielen Fällen wird er aber für die Unterstützung dankbar sein, denn sie kommt der Edition zugute.

Insgesamt stellt die Informatik also einige Mittel bereit, um Editionen heute reicher zu machen, als es mit der gedruckten Form jemals möglich war. Der Arbeitsaufwand wird dadurch eher noch größer, weil ja die ganzen Bausteine erfasst und eingegeben werden müssen. Dafür sind sie dann aber auch in viel größerem Umfang nutzbar. Und wer doch lieber eine gedruckte Fassung in den Händen halten möchte, kann auch diese ohne Probleme erstellen lassen – mit ein bisschen Aufwand (nicht anders als früher, als man ja auch noch setzen musste) auch in einer optisch sehr ansprechenden Form.