

Hochschule Anhalt Fachbereich Informatik und Sprachen

M.Sc. Softwarelokalisierung

Masterarbeit

Erstellung und Lokalisierung von EPUB-Dokumenten

Eingereicht von:

Yaochen Wang

4051638, geboren am 06.05.1987 in China

angestrebter akademischer Grad

Master of Science (M.Sc.)

yaochen.wang@student.inf.hs-anhalt.de

Erstgutachterin: Prof. Dr. Uta Seewald-Heeg

Zweitgutachter: Dipl.-Übers. (FH) Katja Müller

Eingereicht am:

Übersicht

EPUB als das offizielle Standardformat für digitale Bücher gehört heutzutage zu den bedeutendsten E-Book-Formaten der Welt. Die vorliegende Arbeit befasst sich mit der Erstellung von EPUB-Dokumenten und deren Lokalisierung vom Deutschen ins Chinesische am Beispiel des Modulhandbuchs für den Masterstudiengang Softwarelokalisierung an der Hochschule Anhalt.

Diese Arbeit geht davon aus, dass Informationen in der Praxis anstatt in gedruckter Form in elektronischer Form standardisiert angeboten und verwendet werden sollen. Dafür werden die technischen und theoretischen Grundlagen des E-Book-Formats vorgestellt. Vor allem werden die Strukturen und Merkmale von EPUB dargestellt. Zur Ausbreitung des E-Books auf weitere Zielmärkte ist die Lokalisierung solcher Produkte empfehlenswert. Dafür wird aufgezeigt, wie Lokalisierungstechnologien und –werkzeuge helfen, ein EPUB-Dokument zu übersetzen und somit eine bessere Qualität zu gewährleisten. Weiterhin wird die Problematik der Erstellung sowie die Lokalisierung von EPUB-Dokumenten an sich beleuchtet. Zuletzt werden eine Zusammenfassung im Verhältnis zu dem ganzen Projekt und ein Ausblick illustriert.

Abstract

Because of the transform of the mainstream of information media nowadays, it is vital to constitute a standard of electronic publishing in this emerging field. EPUB is the key standard of the organization IDPF, which is the official standard format for digital books. The present dissertation aims at creating and localizing EPUB documents of the module handbook for the master's degree program software localization at Anhalt University of applied Sciences (Hochschule Anhalt) from German into Chinese as example.

This dissertation describes the technical and theoretical foundations of e-book formats and represents especially the structures and features of EPUB. In order to spread the information widely in the world, localization of the products plays an important role. An introduction of localization technology and localization tools is constructed and included in the project. The work processes are mentioned, so as to show, how to translate an EPUB document and ensure good quality. The summary and the view of EPUB are illuminated at the end.

Inhaltsverzeichnis

Abbildungsverzeichnis	VI
Tabellenverzeichnis	VII
Quellcodeverzeichnis	VIII
1. Einleitung	1
1.1 Fragestellung.....	1
1.2 Überblick	3
2. E-Book	5
2.1 Bedeutung von E-Books.....	5
2.1.1 E-Books in der Technischen Dokumentation	6
2.1.2 Besonderheiten von E-Books.....	8
2.2 Geschichte der E-Books	11
2.3 Formate von E-Books.....	12
2.3.1 Text- und Word-Dateien.....	14
2.3.2 HTML und XML.....	14
2.3.3 AZW und Mobipocket.....	15
2.3.4 PDF.....	16
2.3.5 EPUB	17
2.4 Plattformen für E-Books	18
2.4.1 Hardware-Plattformen.....	18
2.4.2 Software-Plattformen	20
3. EPUB	22
3.1 Das EPUB-Format.....	22
3.2 Geschichte von EPUB	23
3.3 Vorteile von EPUB gegenüber PDF.....	24
3.4 EPUB-Standards	25
3.5 EPUB-Dokument-Struktur	26
3.5.1 Dateiformate in einem EPUB-Dokument.....	27
3.5.2 MIME-Type-Datei.....	29
3.5.3 META-INF-Verzeichnis	29
3.5.4 Ressourcen-Dateien	31
3.5.4.1 OPF-Datei	31

3.5.4.2	Inhaltsverzeichnis	36
3.5.4.3	Stilvorlagen	40
3.6	Software für die Erstellung von EPUB-Dokumenten	40
4.	Erstellung eines EPUB-Dokuments am Beispiel der Modulbeschreibungen für den Masterstudiengang Softwarelokalisierung	44
4.1	Internationalisierungsanforderungen	44
4.2	Verwendete Software-Produkte	46
4.3	Vorbereitung der Inhaltsdateien und Funktionsdateien	47
4.3.1	Konvertierung und Vorbereitung von Inhaltsdateien	47
4.3.2	Ermittlung von Inkonsistenten	50
4.3.3	Darstellung und Kodierung von Schriften	58
4.4	Erstellen einer EPUB-Datei	58
4.4.1	MIME-Type-Datei	59
4.4.2	META-INF-Ordner mit container.xml	59
4.4.3	OPS-Verzeichnis	59
4.4.3.1	OPS / Datei content.opf	60
4.4.3.2	OPS / Datei toc.ncx bzw. Datei nav.xhtml	63
4.4.3.3	OPS / Datei images /	69
4.4.3.4	OPS / Datei CSS / Modulandbuch.css	69
4.5	Erstellen eines ZIP-Archivs	71
4.6	Überprüfung	72
5.	Lokalisierung des erzeugten EPUB-Dokuments	77
5.1	Lokalisierungsprojekt	77
5.2	Lokalisierungstechnologie	79
5.3	Lokalisierungswerkzeuge	80
5.4	Durchführung der Lokalisierung	81
5.4.1	Vorbereitung der zu lokalisierenden Quelldateien	81
5.4.2	Konfigurieren des NCX-Dateityps in SDL Trados Studio	82
5.4.3	Vorbereitung der Referenzdateien	86
5.4.4	Übersetzung	89
5.5	Test der lokalisierten Anwendung	92
5.5.1	Qualitätssicherung	92
5.5.2	Erstellen des zielsprachlichen EPUB-Dokuments	93

5.5.3 Verfügbarkeitstest des chinesischen EPUB-Dokuments.....	94
6. Fazit und Ausblick.....	96
Literaturverzeichnis:	97
Anhang.....	100

Abbildungsverzeichnis

Abbildung 1. Umfrage unter 318 Medienunternehmen und Verlagen im Arbeitskreis Elektronisches Publizieren im Börsenverein.....	2
Abbildung 2. EPUB-Dokument-Struktur.....	27
Abbildung 3. EPUB erstellen mithilfe von Silgil.....	42
Abbildung 4. EPUB erstellen mithilfe von Adobe InDesign.....	43
Abbildung 5. GILT.....	45
Abbildung 6. Fehlermeldung bei der Konvertierung von Word zu PDF.....	47
Abbildung 7. Fehlermeldung bei Sonderzeichen in XHTML.....	49
Abbildung 8. Unterschied zwischen Minus-, en/em Gedanken- und Überstrich.....	50
Abbildung 9. Inkonsistenz der Ausgangsmaterialien.....	51
Abbildung 10. Zip-Code im Terminal.....	71
Abbildung 11. Fehlende Informationen auf dem Deckblatt.....	73
Abbildung 12. Deckblatt des EPUB-Dokuments in ibooks.....	74
Abbildung 13. Quer-Anzeige in ibooks.....	75
Abbildung 14. Buchinhalt in Adobe Digital Editions.....	76
Abbildung 15. Auswahl der XML-Basis für NCX-Datei.....	83
Abbildung 16. NCX-Informationen in SDL Trados Studio.....	84
Abbildung 17. Analyse des NCX-Quellcodes.....	85
Abbildung 18. Dateierkennung von NCX in Projekteinstellungen.....	86
Abbildung 19. Hinzufügen neuer Terminologie in SDL Trados Studio 2014.....	88
Abbildung 20. Anlegen eines Translation Memorys in SDL Trados Studio 2014.....	89
Abbildung 21. Problematik der Abkürzungen in der Übersetzung.....	90
Abbildung 22. Ins Chinesische übertragende Literaturnamen.....	91
Abbildung 23. Eigennamen in den originalen Formen.....	92
Abbildung 24. Gefundene Inkonsistenz im Ausgangstext mithilfe von TM.....	93

Tabellenverzeichnis

Tabelle 1. Liste der Referenz-Typen innerhalb von <guide>	35
Tabelle 2. Im Projekt auftretende Codes für HTML-Sonderzeichen	49

Quellcodeverzeichnis

Quellcode 1. Container-Datei	30
Quellcode 2. Beispiel OPF-Datei-Metadatei	32
Quellcode 3. Beispiel OPF-Datei-Manifest	33
Quellcode 4. Beispiel OPF-Datei-Spine	34
Quellcode 5. NCX-Datei-Inhalte-Beispiel	37
Quellcode 6. Beispiel XHTML-Inhaltsverzeichnis-Datei	39
Quellcode 7. Der Kopf von XHTML-Inhaltsdateien	48
Quellcode 8. Inhaltsstruktur von Webseite Modul Effizientes Terminologiemanagement	52
Quellcode 9. Deckblatt mit Stylesheet	56
Quellcode 10. Button in der HTML-Datei	57
Quellcode 11. Container-Datei	59
Quellcode 12. Struktur der OPF-Datei	60
Quellcode 13. Struktur der Inhaltsverzeichnis-Datei von NCX	63
Quellcode 14. XHTML-Inhaltsverzeichnis-Datei-Struktur	67
Quellcode 15. Stylesheet-Datei-Struktur	70

1. Einleitung

1.1 Fragestellung

Bücher sind die Träger der menschlichen Zivilisation und haben viel zu dem Fortschritt der Gesellschaft beigetragen. Neben dem Lernen mit Fachliteratur lesen die Leute im Alltag Zeitungen, Magazine, Romane oder auch Gedichte. Durch die Innovation der Kommunikationstechnologie und der Übertragungsmedien hat die Form der Bücher revolutionäre Veränderungen erlebt. Von den Papierbüchern bis zu den elektronischen Büchern ändern sich die Methoden und Gewohnheiten des Lesens von Menschen. Viele sagen, dass die elektronische Publikation eine neue Revolution ist.¹ Die Innovation der Buchform und die damit einhergehenden Veränderungen vom Lernen und der Arbeit haben unser Leben beeinflusst. Deshalb verdient sie weitere Untersuchungen.

Ein E-Book ist eine neue auf Internettechnologie basierende Publikationsform. Bücher lassen sich heutzutage nicht nur in gedruckter Form, sondern auch in elektronischer Form erzeugen. Die Buchinhalte können digital gespeichert werden. Mit der Verbreitung des Internets und elektronischer Produkte gewöhnen sich die Menschen an die Kommunikationsform von elektronischen Büchern und von digitalen Lesewerkzeugen wie E-Book-Reader. Der allgemeine Trend im Verlagswesen geht zu elektronischen Büchern. Der Online-Buchhandel hat sich im letzten Jahrzehnt weltweit rasant entwickelt (siehe Abbildung 1). Immer mehr Menschen und Unternehmen nehmen an der digitalen Welt teil. Viele der aktuellen Topseller im Online-Buchhandel sind E-Books. Für E-Books werden zahlreiche Formate verwendet, darunter enthalten ist EPUB.

¹ Vgl. T. Low (2011): Verlagsgeplauder – Ebooks – technische Revolution oder teure Sackgasse? Online im Internet: <http://verlorene-werke.blogspot.de/2011/07/verlagsgeplauder-ebooks-technische.html> [09.09.2015]

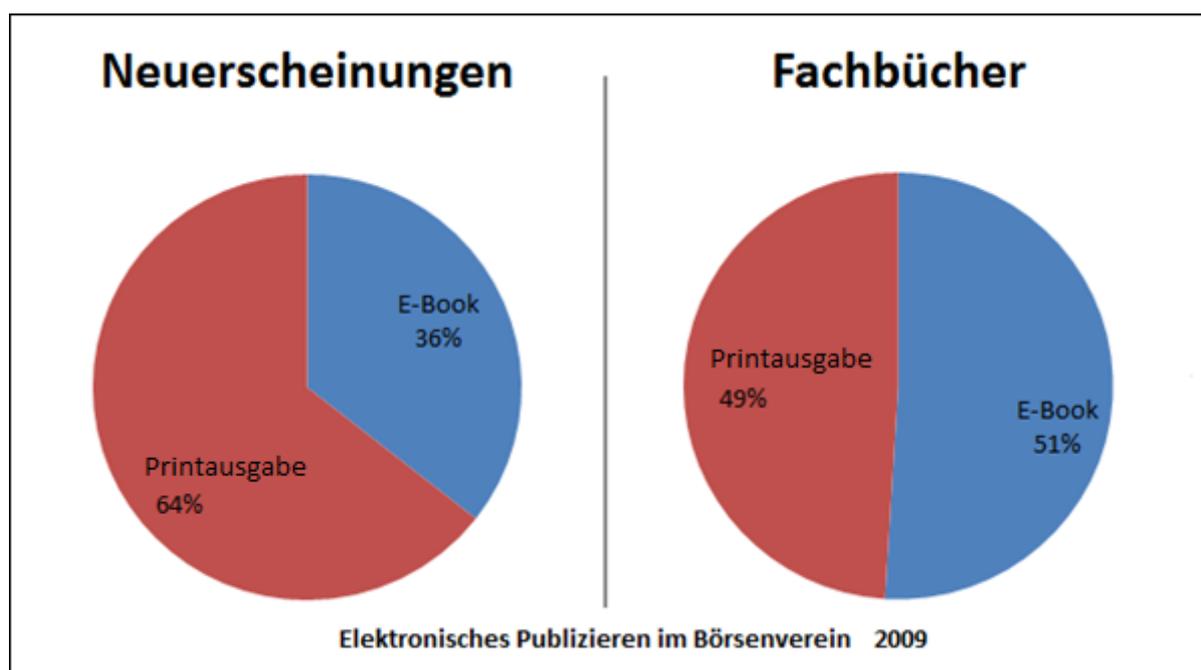


Abbildung 1. Umfrage unter 318 Medienunternehmen und Verlagen im Arbeitskreis Elektronisches Publizieren im Börsenverein²

Nicht nur Bücher, sondern auch andere Publikationen von Informationen werden digital freigegeben, wie zum Beispiel die Handbücher elektrischer Geräte. Für die international tätigen Unternehmen ist eine Vielzahl an Sprachen unter Einfluss der Globalisierung und Internationalisierung zu bedienen, um die Publikationen weltweit auf die Märkte zu bringen. Dabei muss die E-Publikation nicht nur sprachlich präzise, korrekt und verständlich bleiben, sondern auch an den jeweiligen lokalen Markt mit seinen kulturellen Besonderheiten angepasst werden. Hier spielt die fachliche Lokalisierung der E-Publikationen eine wichtige Rolle. Spezifische Lokalisierungstechnologien und Lokalisierungswerkzeuge helfen bei der Reproduktion von E-Publikationen in einer neuen Sprache mit hoher fachlicher Qualität.

Im Mittelpunkt dieser Arbeit steht das für elektronische Publikationen verbreitete Format EPUB. Das Ziel dieser Masterarbeit ist es, die Funktionsweise und die Lokalisierung von EPUB-Dokumenten an einem Beispiel zu zeigen. Dabei wird

² Vgl. Börsenverein des Deutschen Buchhandels (2009): E-Books: Umsatzzweige mit Wachstumspotenzial. Online im Internet: http://www.boersenverein.de/de/158446/Pressemitteilungen/342833?_nav [09.09.2015]

dokumentiert, was man bei der Lokalisierung von EPUB-Dokumenten über EPUB wissen muss. Zunächst wird dargestellt, was ein E-Book ist und welche Vor- und Nachteile E-Books bieten. Ferner wird erläutert, welche Eigenschaften EPUB hat und wie ein EPUB-Dokument erstellt wird. Schließlich wird anhand eines Beispiels dargestellt, wie man ein EPUB-Dokument lokalisiert und worauf man bei der Lokalisierung eines EPUB-Dokuments achten muss. Diese Fragen werden in der vorliegenden Arbeit anhand der Erstellung eines EPUB-Dokuments des Modulhandbuchs für den Masterstudiengang Software-Lokalisierung und dessen Lokalisierung behandelt.

1.2 Überblick

Die vorliegende Arbeit über EPUB-Dokumente und deren Lokalisierung ist in vier Abschnitte untergliedert. Zunächst werden allgemeine Informationen über E-Books dargestellt und die Vor- und Nachteile geschildert. In diesem Zusammenhang werden auch einige wichtige E-Book-Formate und Plattformen von E-Books konkret beschrieben und verglichen. Das Format EPUB wird in Kapitel 2 detailliert dargestellt. Es werden nicht nur der Aufbau eines EPUB-Dokuments illustriert, sondern auch die Funktionen einzelner Dateien eines EPUB-Dokuments charakterisiert.

Wiedergegeben wird auch die manuelle Vorgehensweise der Anfertigung eines EPUB-Dokuments am Beispiel der Modulbeschreibungen für den Masterstudiengang Softwarelokalisierung an der Hochschule Anhalt. Vorausgesetzt ist die Internationalisierungstheorie.

Der Hauptteil der Arbeit beschäftigt sich beispielhaft mit der praktischen Lokalisierung eines EPUB-Dokuments vom Deutschen ins Chinesische, das über die Modulinformationen für den Masterstudiengang Softwarelokalisierung verfügt. In Kapitel 4 werden vor allem die Lokalisierungswerkzeuge und die Lokalisierungsmethode vorgestellt. Zudem werden die fachlichen Lokalisierungsbegriffe und die Funktionsweise der Hilfsmittel geschildert. Im Fokus des Lokalisierungsprojekts stehen spezielle Dateivorbereitungs- und Auslieferungsphasen für die EPUB-Publikation. Dabei werden die Eigenschaften aller

Bestandteile der zu übersetzenden Dokumente beschrieben. Anschließend folgen Hinweise für die Lokalisierung des EPUB-Dokuments. Häufig anzutreffende Probleme in allen Phasen werden auch genannt, analysiert und gelöst. Am Ende ist die Abnahme des chinesischen EPUB-Modulhandbuches.

Zum Schluss werden eine Zusammenfassung über EPUB und die Lokalisierung von EPUB-Dokumenten sowie ein Ausblick in die Zukunft gegeben.

2. E-Book

2.1 Bedeutung von E-Books

Ab Mitte des 20. Jahrhunderts, hat sich die Informatik stark entwickelt. Das Papier ist nicht mehr das einzige Medium zur Speicherung und Verbreitung von Informationen. Menschen haben immer mehr Umgang mit zahlreicheren Informationen und benötigen dafür neue Datenspeicher. So entstanden E-Books, mit denen Daten magnetisch, optisch und elektrisch gespeichert und verbreitet werden können. Das Verlagswesen musste deshalb gewissermaßen verändert werden. Bei der Umwandlung spielten die allmählichen Änderungen der internationalen Lesegewohnheiten und die globalen Anbieter wie Amazon und Google eine bedeutende Rolle.

Ein E-Book ist ein Buch in elektronischer Form. Diese Bezeichnung stammt von den englischen Wörtern „electronic book“. Ins Deutsche übersetzt, heißt das so viel wie E-Buch oder Digitalbuch. Dabei handelt es sich um eine neue E-Publishing-Methode, die durch Einsatz der Online-Technologie geschaffen wurde. Es gibt in der Regel zwei Bedeutungen von elektronischen Büchern. Eine Bedeutung steht für digital in Form einer Softwaredatei gespeicherte Buchinhalte. Manche sind Umwandlungen von echten Büchern und manche sind original digital erstellte Bücher. Die klassischen Aspekte der Papierbücher werden in E-Books mit Hilfe der Computertechnologie simuliert dargestellt und erweitert. Eine andere Bedeutung von E-Books im weiteren Sinne steht auch für elektronische Lesegeräte, anders gesagt E-Book-Reader, wie zum Beispiel E-Papiers und Tablettts. In Abschnitt 2.4.1 wird von E-Book-Readern die Rede sein.

Der Kauf von E-Books ist anders als von Papierbüchern. Dabei handelt es sich meist gleich um den Erwerb von Leselizenzen.³ Die Nutzung mancher E-Books sind durch

³ Vgl. A. Warner(2013): Vom Buch zum Byte. Online im Internet: <http://www.netzpiloten.de/vom-buch-zum-byte-die-geschichte-der-e-books/> [09.09.2015]

die Digitale Rechteverwaltung (DRM)⁴ geregelt. In ähnlicher Weise kann der Autor mit digitalen Wasserzeichen den Missbrauch und die Weiterverbreitung ohne Urheberrecht eines E-Books behindern. Infolgedessen sind das Verleihen an andere Personen, das Erstellen einer Kopie und das Verändern des Buches nicht möglich. Darüber hinaus bieten einige Unternehmen auch Ausleih- bzw. Mietmodelle. Solche E-Books sind nur online für Benutzer zugänglich. Somit hat man „die Möglichkeit, eine bestimmte Anzahl von E-Book-Slots (E-Book-Stellplätze) pro Monat zu erwerben“ [A3W 1012].

2.1.1 E-Books in der Technischen Dokumentation

Als Technische Dokumentation versteht man alle Publikationen zu einem Produkt, die ein technisches Produkt über den gesamten Lebenszyklus beschreiben: Stücklisten, Sicherheitsanweisungen für die Errichtung, Nutzung und das Recycling, CAD-Zeichnungen, und so weiter.⁵ Die Technische Dokumentation umfasst produktbezogene Daten und Informationen, d.h. Unterlagen und Dokumente, die eine Herstellerfirma nach innen oder außen gibt⁶. Technische Dokumentationen werden heutzutage in unterschiedlichen Medien veröffentlicht, um Technikern oder Kunden den freien Zugang zu Informationen zu bieten. So können Technische Dokumentationen auf Papier oder in digitalen Medien verteilt werden.

Umsetzen und Generieren in elektronischer Form ist ein Trend in der Technischen Dokumentation, und zwar die mobile Technische Dokumentation. Als mobile

⁴ DRM als Verkürzung von Digitale Rechteverwaltung bezeichnet Verfahren, mit denen die Nutzung und Verbreitung digitaler Medien kontrolliert werden soll. Sie ermöglicht aus Sicht von Anbietern, die solche DRM-Systeme zur Nutzungskontrolle ihrer Daten einsetzen, um beispielsweise mittels Lizenzen und Berechtigungen sich Nutzungsrechte an Daten vergüten zu lassen. (Vgl. DRM (2015): DRM. Online im Internet: https://de.wikipedia.org/wiki/Digitale_Rechteverwaltung [09.09.2015]) Eigentlich ist DRM als Schutz gedacht, in der Praxis fallen die Einschränkungen häufig in der Nutzung gegenüber Papierbüchern auf. So kann das Lesen auf bestimmte Geräte beschränkt sein.

⁵ Vgl. D. Kirch (2014): Zu Industrie 4.0 durch digitale technische Dokumentation. Online im Internet: <https://www.de.capgemini-consulting.com/blog/digital-transformation-blog/2014/07/eintrag-digitale-technische-dokumentation> [09.09.2015]

⁶ Vgl. Transcom (o.J): Was ist "Technische Dokumentation"? Online im Internet: <http://www.transcom.de/transcom/de/wirueberuns/leistungen/technische-dokumentation.htm> [09.09.2015]

Technische Dokumentation bezeichnet man die jederzeit und überall verfügbare Dokumentation⁷ einer beliebigen Menge. Die Bedienungsanleitungen von Geräten und Anforderungen von Aufträgen werden den Benutzern nicht mehr nur als dicke Bücher, sondern auch als Dateien mitgeliefert. Trotz der Weiterverbreitung der gedruckten Technischen Dokumentation und des Mangels der Integration beider Publikationsmöglichkeiten, werden mobile Technische Dokumentationen heute lieber gelesen.

Inzwischen haben sich vor allem digitale Fachbücher und Handbücher wegen ihrer Suchfunktion und des komfortablen Erstellens von Notizen und Lesezeichen in zahlreichen Branchen durchgesetzt. Texte lassen sich mit Grafiken digital umsetzen. Dabei können Audios, Videos und Verknüpfungen hinzugefügt werden, sodass Methoden oder Anweisungen besser dargestellt werden können. Zudem ist die mobile Technische Dokumentation interaktiver, das heißt Firmen können mit Kunden mittels ergänzender Anwendungen kommunizieren. Wenn es auf eine Vielzahl der Technischen Dokumentationen ankommt, ist die ortsunabhängige Verfügbarkeit der mobilen Technischen Dokumentation von Vorteil. Werden manche Maschinen seit Jahrzehnten noch verwendet, oder ist ein Dokument schwer unter vielen ungeordneten Dateien zu finden, ist es notwendig, die technischen Dokumente zu digitalisieren. Dadurch werden Lagerkosten gespart und ist das Wiederauffinden von bestimmten Informationen aus einer Unmenge von Daten in Sekundenschnelle möglich. Die Einsatzmöglichkeiten im Bereich der Technischen Dokumentation sind als grenzenlos und vielversprechend geschätzt.⁸

Mehr und mehr technische Dokumentationen werden als E-Books angefertigt. Somit kann man die mobile technische Dokumentation auf die heute weit verbreiteten mobilen Endgeräte bequem lesen. Die mobile Technische Dokumentation dient als ein garantierter Vorteil für die Kunden und für langfristige Produktentwicklungen durch günstigere firmenweite Zusammenarbeit. Auf einem mobilen Endgerät können

⁷ Vgl. S. Broda (o.J.): Technische Dokumentation auf mobilen Endgeräten. Online im Internet: <https://www.texniq.de/de/blog/mobile-dokumentation> [09.09.2015]

⁸ Vgl. S. Closs: Dokumentation auf mobilen Endgeräten - Neuauflage für das Single Sourcing. Semcon Communication GmbH. Online im Internet. http://www.doku.info/doku_article_447.html [09.09.2015]

alle Anleitungen eines Haushalts bewahrt und aufgerufen werden. Für Servicetechniker sind Wartungsanleitungen mit Notizen und sogar als Schulungsunterlagen mit Utility Filmen möglich.⁹ Ausreichende Bildschirmgröße, Zoom-/Scroll-Funktion, Navigation und gute Bilddarstellungen sind die Anforderungen eines Lesegerätes für Technische Dokumentation.

2.1.2 Besonderheiten von E-Books

Obwohl E-Books nicht vollkommen sind, bieten sie gegenüber gedruckten Büchern eine Vielzahl an Vorteilen. Unter den folgenden zehn Aspekten werden die Besonderheiten der E-Books beleuchtet:

(1) Verfügbarkeit: E-Books sind schnell zu beschaffen. Wenn die Übertragung zügig ist, das heißt, dass es mit einem schnellen Mobilfunknetz Zugriff auf das Internet gibt, oder Datenaustausch mit Hilfe eines Datenträgers oder eines Kabels möglich ist, lassen sich E-Books jederzeit und allerorts einsetzen. Unabhängig davon, ob sie online gekauft oder kopiert werden, findet eine sekundenschnelle Verbreitung statt, bei der nicht auf die Buchhandlung oder die Post gewartet werden muss. Tausende E-Books lassen sich auf einem E-Book-Reader speichern und so viele auf einmal tragen. Anders als beim normalen Buch, können die meisten E-Books nicht anderen Leuten geschenkt oder auf dem Flohmarkt weiterverkauft werden. Solche E-Books darf man nur selbst benutzen, da das zeitliche Nutzungsrecht der Bücher sich nicht an Dritte übertragen lässt.¹⁰ Jederzeit kann man die virtuelle Buchhandlung aufsuchen und den Bücherbestand aktuell halten.

(2) Zeitlich: Wird der Inhalt eines Buches geändert, kann man seine Version einfach auf die Neuste aktualisieren. E-Books kann man jederzeit verwenden, egal ob auf der Wiese, in der U-Bahn oder am Strand. Bestimmte Wörter oder Seiten sind

⁹ S. Siegel (2010): Technische Dokumentation auf E-Book Readern – geht das? Online im Internet: http://www.tekom.de/uploads/media/208/POD4_Siegel_20813.pdf [09.09.2015]

¹⁰ Vgl. R. Schuldt (2012): eBooks: Was ist erlaubt – und was nicht? Online im Internet: <http://www.computerbild.de/artikel/cb-Ratgeber-Kurse-Internet-eBooks-Was-ist-erlaubt-und-was-nicht-7867389.html> [09.09.2015]

durch die Suchfunktionen schnell zu finden. Beim Lesen kann man die Übersetzungen für Fremdwörter direkt in installierten Wörterbüchern nachschlagen. Jedoch können Festplatten nicht ewig halten und alle Funktionen wegen der technischen Innovationen nicht immer unterstützen, deshalb ist eine Erneuerung von Geräten bzw. Dateiformen nötig.

(3) Benutzertauglichkeit: Man kann die Schriftgröße und -art ändern, dadurch können manche Leute auf die Brille verzichten, um Texte zu lesen. Zudem kann man auch Zeilenabstände, Hintergrundfarbe und Helligkeitsgrad einstellen. Ein E-Book-Reader ist genau so klein und wiegt normalerweise so viel wie ein Taschenbuch, was viel leichter als Fachbücher und Literaturwerke ist. E-Book-Reader erlauben hunderte Bücher ständig überall mit sich herumzutragen. Egal wie viele E-Books ein E-Book-Reader umfasst, ändert sich dessen Gewicht nicht. Eigenschaften von klassischen Büchern, wie Durchblättern und Notizen, werden ohne weitere Werkzeuge wie Stifte in elektrischer Form nachgebildet. E-Books lassen sich außerdem auf verschiedenen mit spezieller Software bzw. Apps ausgerüsteten Geräten verwenden. Durch die Anmeldung desselben Kontos kann man E-Books auf vielen EDV-Geräten¹¹ lesen und eventuell die hinzugefügten Lesezeichen, Markierungen und Anmerkungen synchronisieren. Doch ist der Speicherplatz der aktuellen Geräte begrenzt.

(4) Volumen: Man braucht keinen großen oder zusätzlichen Raum mehr für Bücher und kann dadurch viel Platz sparen. Die Speicher im E-Book-Reader können in der Regel erweitert werden. Sollte mehr Speicherplatz gebraucht werden, kann man einfach ein Speicherchip ins Gerät einsetzen oder den Vorhandenen durch einen Größeren ersetzen. Mit Hilfe der Ergänzungsprogramme kann man E-Books verwalten und in verschiedenen Ordnern nach Sachgebieten oder Zuständen sortieren. Im Mittelpunkt der Technik steht immer die Erweiterung der Akkulaufzeit. Die limitierte Akkulaufzeit ist im Gegensatz zu Papierbüchern eine Beschränkung. Jeder Durchbruch einer längeren Akkulaufzeit ermöglicht einen vielschichtigen Gebrauch des Gerätes.

¹¹ EDV (Elektronische Datenverarbeitung) ist der Sammelbegriff für die Erfassung und Bearbeitung von Daten durch elektronische Maschinen oder Computer. (Vgl. EDV (2015): EDV. Online im Internet: https://de.wikipedia.org/wiki/Elektronische_Datenverarbeitung [09.09.2015])

(5) Kosten: Wegen der Reduktion der Rohstoffkosten sind E-Books zum größten Teil viel billiger als Papierbücher. Auf Grund des einfachen elektronischen Erhaltens von E-Books gibt es dafür fast keine Versandkosten.

(6) Sicherheit: Im Gegensatz zu gedruckten Büchern werden keine Beschädigungen von Papier wie Schmutz, Zerknitterung oder Vergilbung aufkommen. Die E-Book-Reader sind auch nicht leicht zu entzünden. Aber falls die Buchdatei oder das Speichergerät kaputt geht, beispielsweise beim Sturz, kann ein E-Book nicht mehr zugänglich und nicht mehr lesbar sein. Zur Vorbeugung dieses Problems wird deshalb noch eine gute Reserve oder ein Backup-System benötigt.

(7) Lebensdauer: E-Books lassen sich langfristig aufbewahren. Die Zeit hat dabei keinen Einfluss auf die Qualität der E-Books. Dafür wird aber vorausgesetzt, dass das Format, in dem das E-Book vorliegt, bzw. das Gerät, mit dem man E-Books liest, weiterhin unterstützt wird.

(8) Funktionen: Manche simulierte Funktionen wie Suchen wurden schon in den vorhergehenden Abschnitten erwähnt. Außerdem sind interaktive Nutzungen bei E-Books möglich, z.B. Audios, Videos, kleine Spiele, Puzzles usw. Manche E-Books bieten die Funktion des Vorlesens an, wie Kinderbücher und Hörbücher. Manche Abschnitte, Bilder oder Seiten von E-Books kann man auch in andere Anwendungen kopieren, weiterbearbeiten und speichern.

(9) Umweltschonend: 95% der Rohstoffe für die Papierherstellung kommen aus Holz in Form von Holzstoff, Halbzellstoff oder Zellstoff.¹² Die Entstehung von E-Books reduziert den Druck auf die Wälder, weil für E-Book-Dateien keine Bäume gefällt werden müssen und der Holzverbrauch vermindert wird. Wer bestimmte E-Book-Dokumente nicht mehr will, kann diese einfach löschen. Man braucht sich nicht um die bekannten Umweltschäden der Papierherstellung zu sorgen. Zwar benötigen die Erstellung und die Anwendung von E-Books und E-Book-Readern Strom, Internetanschluss, metallische und plastische Werkstoffe usw., wie alle Elektrogeräte,

¹² Vgl. N.N. (o.J): ROHSTOFFE DER PAPIERHERSTELLUNG. Online im Internet: <http://www.bildungserver.com/schulen/papier/rohstoffe.htm> [09.09.2015]

und erzeugen sie zuletzt auch Elektronikschrott. Allerdings werden die meisten Ressourcen bei der Papierbuchherstellung eingesetzt.

(10) Veröffentlichung: E-Books sind für Autoren und Verlage einfacher und kostengünstiger herzustellen. Viele Formate bieten eine Verschlüsselung (DRM) für ihre Dateien. Zusätzlich eröffnen sich mit dem E-Book neue und kostengünstige Vertriebsmöglichkeiten für den verlagsunabhängigen Autor, der damit bei eigener fachlicher Kompetenz auch komplexe Publikationen der Öffentlichkeit verfügbar machen kann [A3W 2012]. Unternehmen müssen auch nicht wegen der Auflagenhöhe oder der Lagerkosten besorgt sein, da nur einige Festplatten für die Lagerung zahlreicher Auflagen in beliebiger Menge benötigt werden.

2.2 Geschichte der E-Books

E-Books haben erst mit der Deklaration eines Mathematik-Studenten im Jahr 1971 an der Universität von Illinois begonnen, welcher zum ersten Mal die Unabhängigkeit eines elektronischen Texts vom Papier formulierte [Warner 2012]. Das im Jahr 1971 gegründete Projekt Gutenberg¹³ bietet seitdem kostenlose digitale Literatur an [A3W 2012]. Vieles von dem, was E-Books heute ausmacht, wurde schon in den 1960er und 1970er Jahren ausprobiert. Manches bleibt bis heute Utopie [Warner 2012]. 1988 wurde das erste verkäufliche elektronische Buch, der Roman *Mona Lisa Overdrive* von William Gibson, veröffentlicht. Danach haben sich verschiedene Institute und Unternehmen daran versucht, Materialien in digitaler Form umzusetzen. Bertelsmann brachte mit dem BEE-Book inmitten der 1990er Nachschlagwerke in einer mit Multimediainhalten angereicherten Fassung auf CD-ROM heraus. Mehrere Verlage experimentierten mit E-Books. [A3W 2012]. Bereits zu Beginn der 1990er brachte Sony den sogenannten Data Discman auf den Markt, danach 2004 das Lesegerät Librie EBR-1000EP, ab 2006 E-Book-Reader wie der iLiad der Firma iRex

¹³ Das Project Gutenberg (PG) ist eine über das Internet zugängliche und von Freiwilligen erstellte digitale Bibliothek. Dabei handelt es sich in den meisten Fällen um Texte von Autoren, die vor mehr als 70 Jahren gestorben sind und deren Werke daher gemeinfrei wurden, also nicht mehr dem Urheberrecht unterliegen. (Vgl. PG (2015): PG. Online im Internet: https://de.wikipedia.org/wiki/Project_Gutenberg [09.09.2015])

Technologies, der Sony Reader sowie die Modelle Hanlin V2 und V8 von Jinke Electronics [Deutschbein 2013].

Das Lesen wird nicht nur mobil, sondern auch vernetzt [Warner 2012]. Die digitale Welt ermögliche generell mehr Menschen einen Zugang zum Lesen [Hauschild 2014]. „Der Markt für E-Books wächst beständig. Inzwischen haben sich vor allem Fachbücher und Handbücher wegen ihrer Suchfunktion und dem komfortablen Erstellen von Notizen und Lesezeichen in zahlreichen Branchen durchgesetzt. Doch auch viele der aktuellen Bestseller erscheinen inzwischen als E-Books. Gerade „Klassiker der Weltliteratur“ kann man „heute über das Internet völlig legal und kostenlos in digitaler Form herunterladen und lesen!“ [ebooknet 2014]. E-Books „rationalisierten Logistik und Bestellwesen radikal und erreichten damit in kurzer Zeit eine marktbeherrschende Stellung“ [A3W 2012]. Die endgültigen Einflüsse auf den Buchmarkt bzw. auf die Lesegewohnheiten der Menschen sind noch nicht klar. Viele Leute können noch nicht auf die Papierbücher verzichten. „Doch der Abschied vom gedruckten Buch hat längst begonnen“, behauptet Ansgar Warner [Warner 2012].

2.3 Formate von E-Books

Metadaten und Codierung erfordern Standards, und das Gleiche gilt für Formate. Das Fehlen von Standards macht die digitale Publikationsbranche chaotisch und behindert zu einem gewissen Maß die Entwicklung der digitalen Publikation. Während der Entstehung von E-Books werden verschiedene Dateiformate für die Konstruktion und das Anzeigen definiert und im Gebrauch gesetzt, obwohl manche ursprünglich für andere Zwecke entworfen wurden. Anders als gedruckte Bücher betragen die bereits existierenden E-Book-Formate eine große Anzahl. E-Books können meist in vielen Dateiformaten geschrieben werden.

Ein E-Book-Format ist eine Datenkodierung, eine Vereinbarung zur Datenstruktur, um zu unterscheiden, in welcher Form die Daten interpretiert werden können. Wie ein Schlüssel für ein Schloss brauchen verschiedene Dateien unterschiedliche Methoden zum Lesen, Anzeigen, Schreiben, Öffnen und sogar zum Laufen. Die

Buntheit der E-Book-Formate steht auch in Wechselwirkung zu der Vielfalt der Endgeräte. Viele Funktionalitäten der Endgeräte und Normen der E-Books werden in der Tat durch die Konkurrenz verschiedener Hersteller erzeugt.

E-Book-Formate unterteilen sich in zwei Darstellungsmodi: Digital optisches Abbild und strukturell funktionales Abbild¹⁴. Digital optisches Abbild ist ein kopierorientierter Ansatz, welcher ein Dokument überall gleich auf den Bildschirmen anzeigt. Diese Formatierung entspricht der WYSIWYG-Methode, damit die Dokumentation im originalen Format ausgedruckt werden kann. Ein strukturell funktionales Abbild orientiert sich an plattformbasierter optimaler Darstellung, sodass es eine bildschirmspezifische Darstellung ermöglicht. Für das digital optische Abbild sind Dokumente lediglich mobil lesbar gemacht und bei einem strukturell funktionalen Abbild bieten Dokumente mit Hilfe einer App die Möglichkeit, das gesamte attraktive Funktionsspektrum des Endgeräts zu verwenden.

Wegen der Erhöhung der Anforderungen von Benutzern und der technischen Anstrengung aller Firmen werden immer mehr auf Realität angepasste Dateiformate generiert und zur Verfügung gestellt. Nachfolgend sind die üblichen Dateiformate von E-Books, einschließlich die herstellerspezifischer Dateiformate: *.txt, *.doc, *.pdf, *.html, *.epub, *.cbz und *.cbr für Bilderbücher, *.lit von Microsoft, *.hlp als Microsoft Help, *.chm (Compiled HTML Help, Compressed HTML Help oder Compiled Help Module(s) von Microsoft), *.mobi als mobipocket von Amazon, weiterhin *.sep, *.ceb, *.exe, *.nlc, *.pdg, *.wdj, *.caj, usw.

Zu den gängigsten Hapterscheinungsformen elektronischer Publikationen zählen PDF, HTML und EPUB. Im Folgenden werden einige wichtige Formate ausgewählt und aufgeführt.

¹⁴ Vgl. E-Book (2015): E-Book. Online im Internet: <https://de.wikipedia.org/wiki/E-Book> [09.09.2015]

2.3.1 Text- und Word-Dateien

TXT- und Microsoft Word-Dokumente sind die einfachsten Formate eines E-Books. Diese Formate erschienen schon lange bevor E-Books entstanden. E-Books in einem solchen Format sind in der Regel nicht für den Verkauf von E-Books bestimmt. Allerdings bieten in China noch viele Autoren bzw. Online-Buchhandlungen zahlreiche E-Books im TXT-Format an. Obwohl diese Formate in China einen großen Anteil am E-Book-Markt besetzen, spielen sie nur eine untergeordnete Rolle in der Welt. Das Format TXT eignet sich nur für wenige E-Books, die keine Formatierungen brauchen und keine Tabellen oder Grafiken enthalten. Die Microsoft Word-Dokumente besetzen dagegen einen niedrigen Anteil am E-Book-Markt und werden deswegen nicht von allen E-Book-Readern unterstützt. Aber Microsoft Word-Dokumente lassen sich einfach in andere Formate konvertieren. Die beiden Formate sind keine formellen Formate von E-Books und können die Anforderungen der Gegenwart nicht erfüllen. Meistens wurden solche E-Books in einer schlechten Qualität erstellt und die Bearbeitung nicht mit dem Einverständnis der Rechteinhaber begonnen. Zusätzlich gibt es noch RTF. Dieses Format kann man als eine Art Sicherheitskopie ansehen, welches benutzt wird, wenn Microsoft Word nicht unterstützt werden.

2.3.2 HTML und XML

E-Books gehören „zur ursprünglichen Konzeption des World Wide Web und seiner Auszeichnungssprache HTML, wofür bestimmte META-Angaben vorgesehen waren, die eine sequentielle Abfolge von HTML-Dateien kennzeichnen sollten“ [A3W 2012]. Ein E-Book im HTML-Format gestaltet mit reinen Texten und besonderen Formatierungen eine einzige Datei als Webseite, die den kompletten Inhalt darstellt. „Die Navigationsmöglichkeiten darin bleiben meist auf die Scroll-Leiste oder simple Kapitelanwahl beschränkt“ [A3W 2012]. Man kann ein E-Book im Format von HTML mit Browser lesen. HTML dient heute als Basis für komplexere E-Book-Formate wie

AZW¹⁵ und EPUB, indem jede Inhaltsdatei im Format von HTML mit anderen Materialien in einem ZIP-Archiv gebündelt wird. Allerdings sind viele Eigenschaften in E-Books im HTML-Format nicht anzutreffen.

XML ist eine Auszeichnungssprache, um Daten zu tragen. In erster Linie dient XML als Datenträger, in dem die Daten vom Format unabhängig sind und durch Style Sheet Sprache wie CSS und XSLT¹⁶ formatiert angezeigt werden können. „Speziell in Russland hat der offene, auf XML basierende Formatstandard FictionBook hohe Verbreitung gefunden“ [A3W 2012].

2.3.3 AZW und Mobipocket

Amazon hat das Lesegerät Kindle produziert, um seine Kunden mit E-Books zu bedienen. Speziell dafür hat Amazon sein eigenes Format, Mobipocket oder kurz Mobi genannt, weiterentwickelt. Mobipocket wurde im März 2000 erfunden und wurde im Jahr 2005 von Amazon aufgekauft. Amazon „setzt bei seinen Kindles auf das sogenannte mobi-Format mit der hauseigenen azw-Verschlüsselung, wodurch der Einkauf von E-Books auf den Shop des Herstellers beschränkt wird“ [Deutschbein o.J.]. Dabei handelt es sich um ein proprietäres Format von Amazon. AZW steht vermutlich für Amazon World. AZW3 ist das primäre Format für Amazon Kindles und hat allmählich das ältere Format Mobipocket ersetzt. AZW4 ist ein proprietäres Format von Amazon für Textbücher. Es ist ein PDF-Format in einer PDB¹⁷-Hülle und meist mit DRM ausgerüstet.¹⁸ Das Mobipocket-Format basiert auf dem HTML-Format und das AZW-Format basiert auf dem Mobipocket-Format. AZW, als Erweiterung von Mobipocket, kann nur auf dem Amazon Kindle oder Softwareprodukten, die mit Kindle kompatibel sind, verwendet werden. Das bedeutet, dass kein anderes Lesegerät E-Books von Amazon ohne Umwandlung öffnen kann.

¹⁵ Vgl. AZW und Mobipocket im Abschnitt 2.3.3.

¹⁶ XSLT ist eXtensible Stylesheet Language Transformations, das empfohlene Style Sheet Sprache für XML. (Vgl. W3S (2015): XSLT. Online im Internet: http://www.w3schools.com/xml/xml_xsl.asp [09.09.2015])

¹⁷ PDB steht für Program Database, ist das Datenbankformat auf PalmOS-System. (Vgl. PDB (2015): PDB Online im Internet: http://wiki.mobilerread.com/wiki/PDB#Palm_Database_Format [09.09.2015])

¹⁸ Vgl. AZW (2015): AZW. Online im Internet: <http://wiki.mobilerread.com/wiki/AZW> [09.09.2015]

¹⁹ Die innere Struktur von AZW3 ist KF8 (Kindle Format Version 8), das das E-Book-Format für den Amazon Kindle Fire ist. AZW3 unterstützt HTML5 und CSS3, die nicht durch Mobipocket unterstützt wurden. Dadurch haben sich die Schwächen des Layouts der Inhalte stark verbessern können.

2.3.4 PDF

Das am weitesten verbreitete und herkömmliche Format für E-Books ist PDF. PDF, die Kurzform von Portable Document Format, wurde von Adobe erfunden und erstmals im Jahr 1993 veröffentlicht. Das Format wurde entwickelt, um ein plattformunabhängiges Mittel für den Dokumentenaustausch von festem Layout zu bieten. Viele Softwareprodukte unterstützen das Erstellen und Anzeigen von PDF, wie zum Beispiel Microsoft Word, Adobe Acrobat und PDFCreator. Allerdings ist PDF kein Standardformat für E-Book-Reader, sondern ein Standard von ISO.²⁰

PDF ist in der Regel für die Druckausgabe konzipiert. Im PDF-Format lassen sich Unterlagen inzwischen auf fast allen mobilen Geräten anzeigen. PDF hat das einfachste und am nächsten an das Seitenlayout eines Buches heranreichende Seitenlayout. Eine PDF-Datei hat wie ein Papierbuch eindeutige Seitenzahlen. Somit können Links und Textverweise im Text beinhaltet bzw. aufgewiesen werden. Dadurch passen PDF-Dokumente positiv zu den Printmedien, Desktop-Bildschirme und Zitierungen. Mit einem festen Layout können PDF-Dateien auf kleinen oder großen Geräten gleich dargestellt werden, ohne eine Veränderung zu erhalten. Der Inhalt und das Layout derselben Datei lassen sich auch nicht ändern. Nicht zuletzt weil PDF-Dokumente leicht von Benutzern auf einer Vielzahl von Computer-Plattformen angesehen und ausgedruckt werden können, werden sie sehr häufig zum Austausch von Dokumenten im Internet verwendet.

¹⁹ Vgl. R. Deutschbein (o.J.): E-Book-Reader als Alternative zum gedruckten Buch. Online im Internet: <http://www.teltarif.de/h/e-book-reader.html> [09.09.2015]

²⁰ ISO ist die Kurzform von International Organisation for Standardization, die Internationale Standards entwickelt und veröffentlicht.

Das fixe Layout ist aber auch ein entscheidender Nachteil von PDF, weil die Inhalte der PDF-Datei sich dadurch nicht gut an die Displaygröße der Endgeräte anpassen lassen. Für Benutzer von ganz unterschiedlichen Endgeräten mit verschiedener Bildschirmgröße ist PDF unter Umständen nicht ganz leserfreundlich. Zum Beispiel ist ein Dokument bei Vergrößerung auf kleinen Netbooks, E-Book-Readern und Smartphones nicht mehr überschaubar. Es ist immer nur ein kleiner Abschnitt sichtbar und man muss deswegen die kleine Anzeige hin und her bewegen, sowie scrollen. Ebenso kann man Abbildungen oder Tabellen auf kleinen Displays nicht gut lesen. Aufgrund dessen sind PDFs nicht dazu geeignet, auf tragbaren mobilen Geräten mit einer kleinen Bildschirmgröße bearbeitet zu werden. Dieses Format passt besser zum Desktop-Computer, denn es erfordert häufig ein Vollpapierformat, welches mit herkömmlichen Bildschirmen leichter dargestellt werden kann.

2.3.5 EPUB

Als Ergebnis zunehmender Einigung auf offene Standardformate, entstand das EPUB-Format, das speziell für E-Book-Reader entworfen ist. EPUB ist ein auf XML basierendes Format zum Anzeigen von E-Books und ist ein quelloffener Standard. Der Inhalt lässt sich automatisch anordnen, das heißt, der Inhalt kann je nach Bildschirmgröße und Plattform angepasst dargestellt werden und daraufhin ein flexibles Seitenlayout anbieten. Die Texte werden heutzutage häufig im sogenannten EPUP-Format bereitgestellt, weil sich der Text dann, je nach Größe, die der Leser individuell einstellt, sich dem Lesefeld anpasst und ständig neu umbricht. Manchmal ist dieser Open-Source-Standard integriert und ein anderes Mal wird eine App zur Ermöglichung der EPUB-Darstellung benötigt. Viele E-Book-Reader, Android-Geräte, PCs usw. unterstützen das Format. Deshalb eignet sich das EPUB-Format für Smartphones und E-Book-Lesegeräte am besten. „EPUB hat starke Unterstützung und wird zurzeit als das vielversprechendste Format neben Mobipocket angesehen“ [A3W 2012]. Auf EPUB wird in Kapitel 3 ausführlich eingehen.

2.4 Plattformen für E-Books

Neben den zahlreichen Formaten ist es wichtig zu wissen, welcher E-Book-Reader und welches Programm ein bestimmtes E-Book-Format unterstützen. E-Books können im Internet in diversen Formaten und für verschiedene Lesegeräte angepasst heruntergeladen und verwendet werden. Wegen der Erhöhung der Anforderungen von Benutzern und der Entwicklung neuer Gerätetechniken werden immer mehr ausgezeichnete Apps geschrieben und neue Leseendgeräte erfunden. Dadurch wird den Kunden ein besseres Leseerlebnis geliefert.

2.4.1 Hardware-Plattformen

E-Books können sowohl auf speziellen Lesegeräten, den E-Book-Readern, als auch auf mit spezieller Software bzw. Apps ausgerüsteten Endgeräten benutzt werden. Ein E-Book als E-Book-Reader ist ein Ausgabegerät oder ein elektronisches Lesegerät, das durch seinen Speicher hunderte von Bücherdateien bereithält und alle Medien, wie Texte sowie Abbildungen, durch sein Display darstellt und Audios sowie Videos durch seine Soundkarte abspielt. Zu den Merkmalen aktueller E-Book-Reader gehören häufig auch integrierte MP3-Player, Bildbetrachter und Browser.²¹ Bei dieser Gelegenheit sind E-Books tragbare handliche elektronische Ausrüstungen.

Uns steht eine Vielzahl mobiler Endgeräte zur Verfügung. Verschiedene Geräte bieten dem Leser andere Leseerlebnisse. Manche davon werden speziell für das Lesen von elektronischen Büchern gestaltet und besitzen meist große Bildschirme sowie integrierte Netzwerkkarte. Solche Lesegeräte haben meist ein sogenanntes E-Ink-Display (elektronische Tinte), mit dem E-Books verfügt und angezeigt werden können. Diese Art von Display bietet nur eine Schwarz-Weiß-Darstellung. Der Text auf dem Display wird nicht mit Durchlicht wie bei einem Monitor erhellt, deswegen wird der Text nicht flimmern und kann ähnlich wie auf Papier dargestellt und gut gelesen werden. Es verbraucht weniger Strom, da dieser fast nur beim Umblättern

²¹ Vgl. R. Deutschbein (o.J.): E-Book-Reader als Alternative zum gedruckten Buch. Online im Internet: <http://www.teltarif.de/h/e-book-reader.html> [09.09.2015]

benötigt wird, d.h. man kann solch ein E-Book-Reader nach dem Aufladen wochenlang verwenden.²² Zusätzlich ist das Lesen bei Sonnenlicht auch möglich. Für den Gebrauch in einer dunklen Umgebung liefern E-Book-Reader mit beleuchtetem Display auch ein gutes Leseerlebnis. Seit 2007 sind über 10 verschiedene Arten des Amazon Kindles herausgegeben wurden, die wie Lesegeräte anderer Firmen vor allem zum Lesen von E-Books ausgelegt²³. Amazon Kindle ist zurzeit das gängigste Lesegerät. Es liegt hauptsächlich an seiner bequemen und einfachen Nutzung. Möchte man lieber Bücher mit visuell naturgetreuen schwarzweißen Effekten, kann man ein E-Ink-Lesegerät kaufen, das Energie spart und weniger Schäden an den Augen hervorruft. Für längeres Lesen sind diese Anzeigen besser geeignet.

Zum Lesen der E-Books stehen neben speziellen E-Book-Readern wie Kindles noch alle PDA-Geräte²⁴, wie Smartphones, Notebooks und Tablet-PCs zur Verfügung. Allerdings sind PCs sowie Laptops unhandlich. Diese E-Book-Reader haben normalerweise ein farbiges LCD-Display, mit den farbigen Bildern und Magazine besser angezeigt werden. Wenn man sich für farbige und interaktivere E-Books interessiert, sollte man sich solche Endgeräte besorgen.

Die elektronischen Lesegeräte sind immer kompakter und leichter geworden. Tablet-PCs, Smartphones und E-Book-Reader akzeptieren für die Anzeige verschiedene gängige Dateiformate. Viele Handys haben schon bei der Auslieferung eine E-Book Funktion integriert. Mit dem Einsatz von E-Book-Reader-Programmen auf diesen Geräten kann man mehr bedienen, als z.B. ein Buch auf dem Handy aufzurufen. Die Speicherchips und die Akkus als Technikdurchbrüche jeder Generation haben sich immer mit mehr Kapazität immens weiterentwickelt. Es gibt zurzeit verschiedene E-Book-Reader mit unterschiedlichen Bildschirmgrößen und –dimensionen, dafür sind auch noch keine üblichen Bildschirmgrößen vorhanden. Außerdem wird der

²² Vgl. S. Siegel (2010): Technische Dokumentation auf E-Book Readern – geht das? Online im Internet. http://www.tekom.de/uploads/media/208/POD4_Siegel_20813.pdf [2015.09.09]

²³ Vgl. Kindle (2015): Kindle. Online im Internet: https://en.wikipedia.org/wiki/Amazon_Kindle#First_generation [09.09.2015]

²⁴ Ein PDA, Kurzform des Englischen Personal Digital Assistant, ist ein kompakter, tragbarer Computer, der neben vielen anderen Programmen hauptsächlich für die persönlichen Kalender-, Adress- und Aufgabenverwaltung benutzt wird.

Bildschirm als eine wichtige Komponente eines E-Book-Readers immer schärfer und kontrastreicher. In Kombination mit einer Tastatur bzw. Softkeys auf dem Display kann man bestimmte Handlungen durchführen. Zudem bieten mobile Endgeräte mithilfe ihrer Grundfunktionen, wie Bewegungssensoren, mehr Möglichkeiten, um die Benutzerfreundlichkeit zu ermöglichen.

2.4.2 Software-Plattformen

Man kann mit spezieller Software bzw. Apps ausgerüstete Smartphones oder Tablet-PCs ebenfalls zum Lesen von E-Books benutzen. Weil E-Books digital in Geräten gespeichert werden, die Licht, Elektrizität und Magnetismus als Medien haben, müssen die Dokumente mittels bestimmter Programme gelesen, kopiert und übertragen werden. Wie die Dokumentation auf mobilen Endgeräten genau aussieht, ist nicht nur davon abhängig, was dokumentiert wird, sondern auch vom Leseprogramm. Ein leistungsfähiges Leseprogramm unterstützt nicht nur verschiedene Lesemodi, wie das Nachtmodus, sondern auch diverse Sprachen.

Für die Darstellung der in TXT oder HTML erstellten Bücher reicht ein Texteditor oder Browser aus. Adobe ist für das Anzeigen von PDF zu verwenden. EPUB-Dokumente erfordern dagegen neue Programme. Zur gleichen Zeit werden viele verschiedene Softwareprodukte zur Unterstützung des Zeigens von E-Books entwickelt. Manche dieser Programme sind kostenpflichtig.

PDF, TXT, Word- und HTML-Dateien werden von normalen E-Book-Readern unterstützt. Bei den meisten E-Book-Readern kann man darüber hinaus auch EPUB finden, das auf XML basierende Format zum Anzeigen. Manchmal ist ein E-Book-Standard integriert, manchmal ist eine App nötig, um zum Beispiel die EPUB-Darstellung zu ermöglichen.²⁵ Für verschiedene Geräte stehen mehrere Apps zur Verfügung. Auf PCs sind Adobe Digital Editions, Calibre, Nook, FbReader, Sigil, manche Web-Browser, Firefox EPUBReader usw. anwendbar. Nook liefert eine

²⁵ Vgl. Semcon Communication GmbH. Dokumentation auf mobilen Endgeräten - Neuauflage für das Single Sourcing. Online im Internet: http://www.doku.info/doku_article_447.html [09.09.2015]

ausgezeichnete Wirkung auf E-Ink-Geräten.²⁶ Das Android-System ist reich an Programmen, die EPUB unterstützen, wie zum Beispiel, Amazons Kindle und FBReader. Viele mit Android verfügbare Programme können auch in Apple-App-Store heruntergeladen werden. Auf IOS-System ist ibooks empfehlenswert, das sich gut für E-Book-Dokumente eignet. Allerdings ist die Darstellung mit verschiedenen Programmen nicht ganz einheitlich.

²⁶ Barnes & Noble Inc. ist das größte US-amerikanische Buchhandelsunternehmen. Darunter operiert BN.com den Online-Vertrieb via Internet, neben dem Vertrieb von traditionellen Büchern, auch für Vertrieb, Logistik und Kundenservice der elektronischen Lesegeräte. Sie bieten sowohl den mobilen Plattformen Nook-Reader-Software, als auch eigene E-Reader. (Vgl. Barnes & Noble. Online im Internet: Wikipedia: https://de.wikipedia.org/wiki/Barnes_%26_Noble [09.09.2015])

3. EPUB

3.1 Das EPUB-Format

Auf Grund der Vielfältigkeit der E-Book-Formate müssen verschiedene Lese-Programme für unterschiedliche proprietäre und generische Dokumentformate heruntergeladen und auf dem Computer bzw. Endgerät installiert werden, um alle Sorten von E-Books zu lesen. Das ist zum einen "unbequem" und zum anderen eine Verschwendung von Ressourcen, welche nicht nur die Kosten des Lesens für den Nutzer erhöhen, sondern gleichzeitig auch zum kompletten Verlust des Benutzers führen kann, da dieser aufgrund der Komplexität die Freude am Lesen verliert. Mithilfe der Erfahrung des traditionellen Verlagswesens wurde eine Reihe von freien Industriestandards für E-Books etabliert, die vorteilhaft sowohl für das Lesen und das Anwenden von Publikum, als auch für die Entwicklung der elektronischen Publikation sind.

EPUB ist das Akronym für das englische Wort *electronic publication*. Der Standard EPUB ist ein offener und freier Standard für E-Books, das von IDPF(International Digital Publishing Forum) skizziert wurde. Das IDPF ist eine Welthandels- und Standardisierungsorganisation, die sich für die Entwicklung und die Verbreitung des elektronischen Publizierens und des Inhaltskonsums einsetzt.²⁷ Der Haupt-IDPF-Standard ist EPUB. Die IDPF entwickelt und pflegt den Veröffentlichungsstandard von EPUB-Inhalt. Die Arbeit der IDPF fördert die Entwicklung von elektronischen Veröffentlichungsanwendungen und Produkten, die Schöpfer der Inhalte, Hersteller von Lesesystemen und Verbraucher begünstigen. Der Standard ermöglicht die Erzeugung und den Transport der zurückfließbaren²⁸ digitalen Bücher und andere Arten von Inhalten als digitale Veröffentlichungen.

²⁷ Vgl. Idpf (2011): EPUB. Online im Internet: <http://idpf.org/epub> [09.09.2015]

²⁸ Zurückfließbar (Englisch: reflowable) bedeutet, dass sich ein EPUB-Dokument automatisch zu einer Anzeige anpassen und verstellen lässt.

Die EPUB-Spezifikation ist ein Vertriebs- und Austauschformatstandard für digitale Publikationen und Dokumenten. EPUB definiert ein Mittel zur Darstellung, Verpackung und Entschlüsselung strukturierter und semantisch erweiterter Web-Inhalte, darunter HTML5, CSS, SVG, Bilder und andere Ressourcen für die Verteilung in einer einzelnen Datei. EPUB 3.0.1 ist die aktuellste Version des EPUB-Standards.²⁹

E-Books in der EPUB-Form hat die von IDPF etablierte OCF-Spezifikation eingehalten, welche der ZIP-Komprimierungstechnologie folgt. Das heißt, dass ein EPUB-Dokument eine ZIP-Datei ist. Wird die Dateierweiterung .epub in .zip umgeschrieben, kann man ein EPUB-Dokument mit Komprimierungssoftware (wie *WinRAR* oder *WinZIP*) entpacken und durchschauen. Als eine einfache ZIP-Datei, enthält ein EPUB-Dokument geordnete Dateien. Öffnet man mit WinRAR ein EPUB, kann man die Verzeichnisstruktur mit den einzelnen Dokumenten einsehen, die zumeist in XML dargestellt werden.

3.2 Geschichte von EPUB

Zwei Hersteller von E-Book-Systemen (hauptsächlich SoftBook Press und NuvoMedia Inc.) haben „Open eBook“ für ihre mobilen Lesegeräte entwickelt und in den späten 1990er Jahren veröffentlicht. Die Version 1.0 der Spezifikation, nämlich OEBPS (**O**pen **e**Book **P**ublication **S**tructure) bzw. die Kurzform OEB (**O**pen **e**Book), wurde 1998 veröffentlicht.³⁰ Der Standard EPUB hat den älteren Standard OEBPS ersetzt. Gleichzeitig hat die Dateierweiterung .epub die Dateiendung .opf substituiert. Der Vorgänger von IDPF war OeBF (**O**pen **e**Book **F**orum). Der Standard EPUB wurde als eine offene Spezifikation von OeBF im Jahr 1999 eingeführt, um den Inhalt und die Struktur von E-Books zu zeigen und die Vereinheitlichung zu erreichen und Kompatibilität der Formate zu erhöhen. „Das OeBF hat sich das Ziel gesetzt, [OEBPS als] eine einheitliche Plattform für alle E-Book-Anwendungen zu entwickeln

²⁹ Vgl. Idpf (2011): EPUB. Online im Internet: <http://idpf.org/epub> [09.09.2015]

³⁰ Vgl. EPUB (2015): EPUB Geschichte. Online im Internet: <https://de.wikipedia.org/wiki/EPUB> [09.09.2015]

und standardisieren.“ „Zu den Gründern dieser Initiative gehören Microsoft, Palm und Adobe aber auch Unternehmen der Publikationsindustrie.“ [ITWissen 2015] Das ermöglicht diese plattformübergreifende Anwendung von E-Books. Damals war der OCF-Standard noch nicht skizziert. OCF, das im Jahr 2007 entwickelt wurde, definiert die Regeln dafür, wie eine abstrakte Sammlung von Dateien in die physische Darstellung in einem ZIP-Archiv darzustellen ist.³¹ Daher gab es viele Dateien in einem OEBPF-Dokument. Auf dieser Basis wurden Änderungen vorgenommen und schließlich entstand das EPUB-Format. Zu bemerken ist, dass EPUB zunächst von dem traditionellen Verlagswesen übernommen worden ist.

EPUB 3.0 hat EPUB 2.0.1 im Oktober 2011 ersetzt. Im Juli 2014 wurde die neueste Version von EPUB 3.0.1 veröffentlicht. Das Konzept von Version 2 ist etwas allgemeiner gehalten als das von Version 3. Die Möglichkeiten von Version 2 wurden teilweise in Version 3 gestrichen. Deshalb bietet Version 3 hinsichtlich der semantischen Textauszeichnung flexiblere Möglichkeiten.

Nach Wikipedia³² kann „ab Version 3.0.1 zudem festgelegt werden, ob die Präsentation seitenbasiert oder rollbar sein soll. Zugleich kann festgelegt werden, wie sich der Inhalt an den verfügbaren Platz anpasst – entweder [...] durch Festlegung eines Darstellungsbereiches und gegebenenfalls anschließender Skalierung ('fixiert') oder [...] durch automatische Aufteilung des Inhaltes auf den verfügbaren Platz ohne Skalierung“ [EPUB3 2015].

3.3 Vorteile von EPUB gegenüber PDF

PDF ist eine Datei, während ein EPUB-Dokument eine Anwendung ist. Beide Formate haben ihre Vor- und Nachteile.

Die PDF-Datei ist plattformunabhängig. PDF eignet sich besser für Fachliteratur, in der viele Grafiken, Tabellen oder Formeln eingebunden sind. Die als EPUB erstellten

³¹ Vgl. Idpf (2011): EPUB. Online im Internet: <http://idpf.org/epub> [09.09.2015]

³² ebd.

Dateien passen sich hingegen sehr gut zur Bildschirmgröße von Lesegeräten an, insbesondere zur Darstellung von Fließtext auf kleinen Geräten. Für Belletristik oder Sachbücher ohne viele Grafiken oder Tabellen ist EPUB daher ein gut geeignetes Format. Das Lesen von PDF ist allerdings nur auf Geräten mit größerem Display wie Computer oder Tablets zu empfehlen.³³ Das EPUB-Format eignet sich insbesondere für die Ausgabe auf kleinen bis mittleren Geräten. Demzufolge erlauben auf EPUB basierende E-Books eine dynamische Anpassung an das jeweilige Gerät und die auf PDF basierenden E-Books immer genauso statisch formatierte Darstellungen. Außerdem ist PDF keine einfache Sprache und schwer zu erlernen. Bei EPUB sind nur grundlegende Kenntnisse für HTML und XML ausreichend. Zudem ist EPUB nicht nur für Bücher zu verwenden, sondern auch für Verpackungen von Dokumenten, Blogs und anderen Web-Informationen zu nutzen.

3.4 EPUB-Standards

EPUB ist eine „Verpackung“ von freien Formatstandards. Unter dem Namen EPUB fasst IDPF hauptsächlich drei offizielle Substandards zusammen:³⁴

OPF (Open Packaging Format)-Standard kümmert sich um den Aufbau eines EPUB-Dokuments. Dieser Substandard von EPUB beschreibt die Struktur der EPUB-Dateien in XML. OPF enthält die Metadaten sowie benötigte Dateien der Publikation und legt die Reihenfolge der Inhalte fest. Der erweiterte Navigationsstandard ermöglicht direkten und leichten Zugriff auf einzelne Inhalte (entsprechende Datei *content.opf*, *stylesheet.css* bzw. *XHTML-Inhaltsverzeichnisdatei*);

OPS (Open Publication Structure)-Standard bezieht sich auf das Design der eigentlichen Inhaltsdaten. Dieser Standard beschreibt die Struktur der eigentlichen Inhalte und die Darstellung der Publikation, damit die

³³ Vgl. N.N. (o.J.): EPUB oder PDF - welches Dateiformat ist besser? Häufig gestellte Fragen - FAQ zu eBooks. Online im Internet: <http://www.lehmanns.de/page/ebookfaq> [09.09.2015]

³⁴ Vgl. Idpf (2011): EPUB. Online im Internet: <http://idpf.org/epub> [09.09.2015]

Darstellungsäquivalenz auf verschiedenen Lesegeräten ermöglicht (entsprechende alle Dateien);

OCF (Open Container Format)-Standard, das von OEBPS Container Format weiterentwickelt ist, definiert das Einzeldateiformat, die Verzeichnisstruktur und Dokumentstruktur der Publikation. Das OCF stellt eine so genannte Container-Technologie dar. Als ein Behälterformat verpackt OCF alle Daten in ein Endformat, als eine Einzeldatei, indem es alle benötigten Dateien in eine ZIP-Datei kapselt. Diese Technologie ermöglicht das leichte Verteilen und der Vertrieb sowie inhaltsverlustfreien Transport eines Dokuments. Durch eine klare Dateistruktur ist auch das Datenmanagement und Zugriff auf einzelne Dateien möglich. Das OCF kann zusätzlich XML-Technologien einbinden, die digitale Signaturen von Wasserzeichen oder Verschlüsselungsmechanismen enthalten. Zudem hat das OCF noch grundlegende Formate wie zum Beispiel die Kodierung von Dateinamen und die Beschränkung des Zip-Standards spezifiziert. (entsprechende Datei *mimetype.xml* und Verzeichnis *META-INF*)

Wie bereits im Abschnitt OCF erörtert, werden im EPUB-Standard für manche bestimmte Formate, für andere einzelne Standards eingeschlossen, zum Beispiel XML, XHTML, CSS, ZIP usw. Weiterhin werden in Version 3 noch Medienüberlagerungen (für eine alternative Präsentation als Hörbuch) und kanonische Fragmentidentifizierer (für Verweise auf beliebige Stellen in einem Buch) zugegeben. Auf manche konkrete Spezifikationen wird in folgenden Abschnitten noch detailliert eingegangen.

3.5 EPUB-Dokument-Struktur

Ein unverschlüsseltes EPUB-Dokument lässt sich zumeist in drei Teile gliedern und schließt dabei folgende Dateien ein:

1. Datei *mimetype.xml*
2. Verzeichnis META-INF, darin Datei *container.xml*

3. Verzeichnis OPS oder unter anderem Namen, darunter *OPF-Datei*, XHTML-Dateien mit Inhalte und Titel, NCX- oder XHTML-Navigationsdatei, eventuell Format-Datei *stylesheet.css* sowie Image-Verzeichnis, darunter die einbezogene Grafiken. Weitere audiovisuelle Referenzen sind auch möglich. (siehe Abbildung 2)

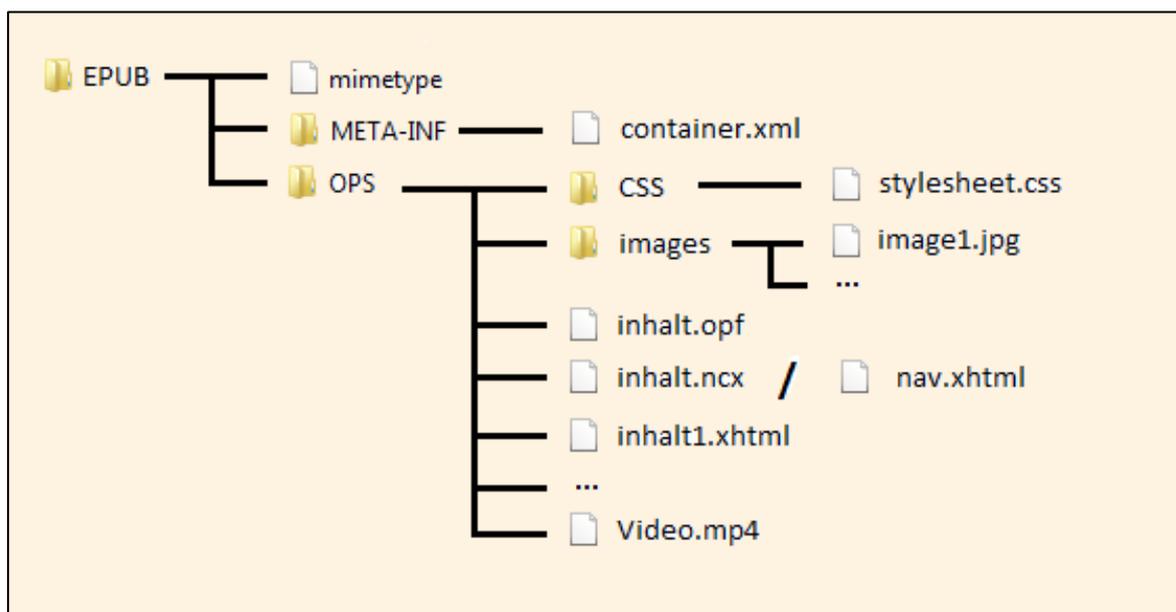


Abbildung 2. EPUB-Dokument-Struktur³⁵

3.5.1 Dateiformate in einem EPUB-Dokument³⁶

Diese Spezifikation EPUB erkennt die Bedeutung der aktuellen Software-Tools, Altdaten, Veröffentlichungspraktiken und Marktbedingungen. Daher hat EPUB die bevorzugten Regelungen eingesetzt.

Inhalte des EPUB-Dokuments sind im Allgemeinen (X)HTML-Dateien³⁷. Und deren Beschreibungen liegen alle als XML-Dateien vor.³⁸ Außerdem erscheint noch die

³⁵ Eigene Darstellung.

³⁶ Vgl. IDPF. (2015): EPUB. Online im Internet: <http://idpf.org/epub> [09.09.2015]

³⁷ Die Inhaltsdokumente können ab EPUB Version 3 von HTML 5 realisiert werden.

³⁸ Als Alternative bietet DTB (Digital Talking Book) von DAISY (Digital Accessible Information System) eine semantisch deutlicher Struktur für Inhalte an. Weil DTB weniger als XHTML bekannt und

CSS-Format-Datei, die ganz anders formuliert ist. Mit Hilfe von CSS kann die Darstellung von (X)HTML festgelegt werden. EPUB Version 3.0.1 unterstützt Medienformate und Pixelgraphikformate: PNG, JPEG/JFIF, GIF und SVG³⁹.

OPS und OPF basieren auf XML, wegen sowohl seiner Allgemeinheit und Einfachheit, als auch der Wahrscheinlichkeit der zukünftigen Anpassung an die Technologien und des Gebrauchs. XML bietet auch klar definierte Regeln für die Syntax von Dokumenten, die die Kosten und Inkompatibilität verringert. Weiterhin ist XML erweiterbar, d.h. XML unterstützt die Internationalisierung und Auszeichnung.

Neben den Konformitätskriterien muss ein OCF-Abstrakt-Behälter zusätzlich die folgenden Bedingungen erfüllen:⁴⁰

- Alle XML-Dateien müssen gut ausgebildet sein (wie in XML 1.0 definiert) und umfassen somit eine korrekte XML-Deklaration (z.B. `<? Xml version = "1.0"?>`).
- Alle XML-Dateien müssen mit der XML 1.0 und dem Namensraum in der XML-Spezifikation kompatibel sein; Alle XML-Dateien müssen in UTF-8 oder UTF-16 codiert werden.
- Alle XML-Dateien müssen jeweiliger angegebener MIME-Type der relevanten XML-Spezifikation entsprechen.

Jedoch werden nicht alle XHTML 1.1 Elemente und Attribute⁴¹ in dieser Spezifikation enthalten.

gebraucht ist, zudem ist DTB kaum von Programmen unterstützt, wird in dieser Arbeit nicht näher auf diese Variante eingegangen. Bei EPUB 3 ist es wegen der amgelnden Interpretation bereits gestrichen. Wikipedia online im Internet.

³⁹ SVG (Scale Vector Graphics) ist ein internationaler Satandard zur Beschreibung zweidimensionaler Vektorgrafiken.

⁴⁰ Vgl. IDPF. (2010): Open Container Format (OCF) 2.0.1. Online im Internet: http://www.idpf.org/doc_library/epub/OCF_2.0.1_draft.doc [09.09.2015]

⁴¹ Attribute sind die Eigenschaften eines Elements. Sie können ein Element genauer spezifizieren und normalerweise in einer Attributlisten-Deklaration vordefiniert.

3.5.2 MIME-Type-Datei

MIME-Type bezeichnet die Internet Media Type. MIME steht für *Multipurpose Internet Mail Extensions* und definiert ursprünglich das Datenformat von E-Mails. Durch die Entwicklung seines Schemas können der interpretierenden Software den Datentyp des nächsten Teils der Email mitgeteilt werden. Das Schema gilt auch für die zu übertragenden Daten im gesamten Internet.⁴²

Die im Web-Browser angezeigten Inhalte werden in unterschiedlichen Formen dargestellt, zum Beispiel HTML, XML, GIF oder Flash. Weil im HTML Hyperlinks enthalten sein können, sind jegliche Mediendateien von HTML abrufbar. Der Browser entscheidet durch die MIME-Type, und zwar die Internet-Media-Type, wie die Daten anzuzeigen sind. Die Internet-Media-Type wird einem Browser normalerweise durch HTTP-Übertragung vom Webserver mitgeteilt. Jeder Internet-Media-Type besteht aus zwei Teilen: Einerseits die Angabe eines Medientyps von Daten, wie zum Beispiel *image*. Andererseits die Angabe eines Subtyps, wie zum Beispiel *jpeg*; zusammen geschrieben *image/jpeg*.

Jedes EPUB-Buch enthält eine MIME-Type-Datei mit dem gleichen Inhalt: „application/epub+zip“, dabei sind Umbrüche und neue Zeilen nicht erlaubt. Diese Information gibt das Dateiformat des E-Books an und zeigt, dass sich diese Datei mit einem E-Book-Reader bzw. ZIP-Werkzeug öffnen lässt. MIME-Type ist die erste Datei in der EPUB-Verpackung und muss unkomprimiert im Stammverzeichnis bleiben.

3.5.3 META-INF-Verzeichnis

Das Verzeichnis META-INF muss im Stammverzeichnis eines EPUB-Dokuments liegen und darin eine Container-Datei enthalten, die beständig vorhanden sein muss. Als eine XML-Datei soll die Container-Datei mit <?xml>-Prolog beginnen. Der

⁴² Vgl. MIME-Typen (2015): Referenz: MIME-Typen. Online im Internet: <https://wiki.selfhtml.org/wiki/Referenz:MIME-Typen> [09.09.2015]

Namensraum des Wurzelements zeigt, dass das Konzept dem Open Document Format entlehnt wurde.⁴³ Ein EPUB-Reader wird vor allem diese Container-Datei lesen, die zum einen auf das Wurzelverzeichnis und die OPF-Datei mit Metadaten verweist und zum anderen die Angabe über die Möglichkeit, das Dokument zu öffnen, definiert, indem der Medientyp unter dem Attribut *media-type* angegeben ist. Das Element `<rootfile>` kann mehr als einmal genutzt werden, um sicherzustellen, dass ein Reader das Dokument in einem von allen angegebenen Formaten darstellen kann, sollte das erste Format nicht unterstützt werden. Das erste `rootfile`-Element in der Container-Datei verweist auf die EPUB-Variante und gibt dessen Medientyp "application/oebps-package+xml" an. Im Folgenden steht ein Beispiel der Inhalt einer Container-Datei (Siehe Quellcode 1).

Quellcode 1. Container-Datei⁴⁴

```
<?xml version="1.0" encoding="UTF-8"?>
<container xmlns="urn:oasis:names:tc:opendocument:xmlns:container" version="1.0">
  <rootfiles>
    <rootfile full-path="OPS-Ordner/content.opf"
      media-type="application/oebps-package+xml"/>
  </rootfiles>
</container>
```

Neben der immer vorhandenen Container-Datei sind noch einige Dateien im EPUB-Dokument optional:

- manifest.xml: Manifest-Datei für Nicht-OPS-Inhalte
- metadata.xml: Metadaten für Nicht-OPS-Inhalte
- signatures.xml: Digitale Wasserzeichen
- encryption.xml: Datenverschlüsselung
- rights.xml: Rechtemanagement

⁴³ Vgl. Data2type (o.J.): ePUB. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

⁴⁴ Ebd.

Die *manifest.xml* und *metadata.xml* sind nur dann nötig, falls es neben OPS-Inhalte noch weitere Dateien gibt.⁴⁵

3.5.4 Ressourcen-Dateien

Alle Ressourcen-Dateien wie eingebettete Medien müssen im EPUB-Archiv liegen. Dabei sind externe Bestände nicht erlaubt. Das OPS-Verzeichnis ist ein Unterverzeichnis im EPUB-Dokument, das sich umbenennen lässt. Das Verzeichnis dient zur Speicherung aller Inhaltsdateien und Stilvorlagen sowie Unterverzeichnissen für Abbildungen usw. Darunter sind die Dateien *content.opf* und *toc.ncx* zwangsläufig und andere je nachdem optional. Die voraussichtlichen Dokumente können OPF, CSS-, NCX- und TTF-Dateien sein. Die meisten Dateien kann man umbenennen, doch die Dateierweiterungen dürfen in der Regel nicht geändert werden.

3.5.4.1 OPF-Datei

Die OPF-Datei als eine normative XML-Datei ist die Kerndatei von EPUB. Die Aufnahme von Metadaten und die Strukturierung werden durch OPF geregelt. Eine OPF-Datei muss den Namensraum „<http://www.idpf.org/2007/opf>“ enthalten und zur Angabe weiterer Metadaten zusätzlich den Namensraum der Dublin Core Metadata Initiative (DCMI) „<http://purl.org/dc/elements/1.1/>“ besitzen. Aber Dublin Core gehört nicht zu EPUB. Eine OPF-Datei verweist auf alle Bestandteile der elektronischen Publikation. Neben den Eingangsdaten (Definitionen von DOCTYPE und Verpackung) besteht OPF hauptsächlich aus fünf Blockangaben: `<metadata>`, `<manifest>`, `<spine>`, `<guide>` und `<tour>`. Alle Informationen sind mit standardisierten Elementen strukturiert oder ausgezeichnet. Die entsprechenden Elementnamen tragen alle das Präfix *dc* des Dublin-Core-Namensraums.

⁴⁵ Siehe Fußnote 42

Der erste Block <metadata> bietet Metainformationen der Publikation an und wird gemäß den 15 Dublin-Core-Elementen (engl. Dublin Core Elements) vom Dublin-Core-Schema bereitgestellt. Die Elemente <dc:title>, <dc:Identifizier> und <dc:language> sind zwingend erforderlich. Andere sind optional, können mehrfach auftauchen und in beliebigen Reihenfolgen stehen. Darin handeln <dc:subject>, <dc:coverage> und <dc:description> von der Beschreibung des Inhalts. Die Felder <dc:format> und <dc:type> beziehen sich auf die Technischen Daten. Außerdem gehen die Felder <dc:creator>, <dc:publisher>, <dc:contributor>, <dc:rights>, <dc:source>, <dc:relation> und <dc:date> um die Veröffentlichungsdaten von Personen und Dokumentangaben. Das Element <dc:identifizier> ist die eindeutige Identifizierung des Dokuments, damit sich das Dokument zitieren lässt. Der Wert von <identifizier> darf vom Ersteller definiert werden, muss aber einzigartig sein. Für den Buchverlag muss in dem Feld die ISBN oder die Nummerierung der Library of Congress enthalten sein. Die Angabe des Attribut *ids* muss mit dem Attribut von „unique-identifizier“ in den Eingangsdaten übereinstimmen. Das Element <dc:language> legt die Sprache der EPUB-Publikation fest, dessen Werte dem unter XML/XHTML genannten Sprachattribut *xml:lang* entspricht.⁴⁶ Um die beschreibenden Informationen genauer zu definieren, können die Tags noch mit OPF-Attributen gemäß der Dublin-Core-Grammatik erweitert werden. Zusätzliche Metadaten, die in Dublin Core nicht existieren, können in XHTML-Syntax mit Hilfe des Elements <meta> formuliert werden. Nachfolgend ist ein Beispiel des Inhalts vom Element <metadata> einer OPF-Datei (Siehe Quellcode 2):

Quellcode 2. Beispiel OPF-Datei-Metadaten⁴⁷

```
<metadata xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>Titel</dc:title>
  <dc:identifizier id="PrimaryID">ID</dc:identifizier>
  <dc:language>Sprache</dc:language>
  <dc:date>Datum</dc:date>
  <dc:subject>Genre</dc:subject>
```

⁴⁶ ISO 639-1 definiert Abkürzungen für Sprachen. In HTML und XHTML können sie bei den Attribute *lang* und *xml:lang* verwendet werden. Deutsch als “de”, Chinesisch als “zh” und Englisch als “en”. Vgl. w3schools. Online im Internet: http://www.w3schools.com/xml/xml_xsl.asp [09.09.2015]

⁴⁷ Vgl. Data2type (o.J.): Die Metadaten der Publikation. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

```
</metadata>
```

Im Abschnitt `<manifest>` müssen alle im Archiv gespeicherten Dateien aufgelistet werden. Darin muss mindestens ein Unterelement vorhanden sein. Jedes Unterelement bezieht sich auf eine Komponente, die ein Teil vom EPUB aufbaut. Diese Aufgabe übernimmt das Unterelement `<item>`. Alle Unterelemente müssen die Attribute `id`, `href` und `mediatype` haben. Die Reihenfolge aller Elemente ist hier nicht von Bedeutung. Doch müssen alle Inhaltsdateien, Dateien für Stilvorlagen und Grafiken aufgelistet sein. Quellcode 3 zeigt die möglichen Elemente in `<manifest>`.

Quellcode 3. Beispiel OPF-Datei-Manifest⁴⁸

```
<manifest>
  <item id="inhalt1" href="inhalt1.xhtml" media-type="application/xhtml+xml"/>
  <item id="inhalt2" href="inhalt2.xhtml" media-type="application/xhtml+xml"/>
  <item id="main-css" href="css/stylesheet.css" media-type="text/css"/>
  <item id="ncx" href="content.ncx" media-type="application/x-dtbncx+xml"/>
</manifest>
```

Danach kommt das Element `<spine>`. Das englische Wort „spine“ beschreibt das "Rückgrat" eines Buches, nämlich der Buchrücken. `<spine>` als ein minimales Inhaltsverzeichnis beinhaltet nur ein Unterelement `<itemref>`, indem alle Inhaltsdokumente enthalten sollen.

Die Inhaltsdokumente entsprechenden XHTML-Dateien, mit denen ein E-Book zusammengestellt wird, und anderen relevanten Dateien, die in der Navigation benötigt sind. Jedes `<itemref>` verweist auf ein im `<manifest>` bezeichnetes Inhaltsdokument. Das referenzierende Attribut `idref` richtet sich gegen den `id`-Wert eines Elements von `<item>` innerhalb von `<manifest>` und ist daher ein Pflichtattribut.

Die Reihenfolge der Elemente legt die zugehörigen Inhaltsdokumente in die lineare Lesereihenfolge fest. Die geordnete Liste ist die Grundlage sowohl für die Inhaltsverzeichnis-Datei als auch für die Elemente `<tours>` und `<guide>`. Das

⁴⁸ Vgl. Data2type (o.J.): ePUB. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

Element `<spine>` muss demnach das Attribut `toc` (table of content) einschließen, dessen Wert mit dem ID-Eintrag des erforderlichen NCX-Datei oder XHTML-Inhaltsverzeichnis-Datei (Siehe Abschnitt 4.4.3) im `<manifest>` deklariert ist. Mit dem Attribut „`toc`“ wird auf ein ausführliches Inhaltsverzeichnis verwiesen. Zu weiteren häufig verwendeten Attribut-Werten zählen *loi* für list of illustrations und *lot* für list of tables. Das optionale Attribut *linear* zeigt, ob der Inhalt primär oder sekundär ist. Der Defaultwert „`yes`“ deutet hin, dass der Abschnitt primärer Inhalt ist. Als Defaultwert kann das Attribut bei der Formulierung weggelassen werden. Die mit Linear-Wert „`no`“ gekennzeichneten sekundären Dateien werden von Lesegeräten dagegen im Lesevorgang übersprungen. (Siehe Quellcode 4)

Quellcode 4. Beispiel OPF-Datei-Spine⁴⁹

```
<spine toc="ncx">
  <itemref idref="Inhaltsverzeichnis"/>
  <itemref idref="inhalt1" linear="yes"/>
  <itemref idref="inhalt2" linear="no"/>
</spine>
```

Ein Autor könnte speziell ausgewählte Inhalte eines E-Books in einer Reihe versammeln, um eine besondere Navigation anzubieten. Das sogenannte Element `<tour>` und seine Funktion werden grundsätzlich durch das Navigationskonzept der NCX-Datei ersetzt. Eine nähere Bezeichnung bietet das Führungselement `<guide>`, das grundlegende strukturelle Bestandteile von EPUB identifiziert, um Lesesysteme einen bequemen Zugang zu ermöglichen. Die strukturellen Bestandteile sind beispielsweise das Deckblatt, das Inhaltsverzeichnis, das Literaturverzeichnis u.ä. Das Unterelement `<reference>` von `<guide>` verweist mit zwei Pflichtattributen *href* und *type* auf die Inhaltsdatei und deren Funktion. Nachfolgend ist eine List von möglichen Typen (Tabelle 1). Datensätze von nicht in der List angegebenen Typen können bei Bedarf mit „`other`“ beschrieben werden. Beide, `<tour>` und `<guide>` sind in OPF optional. Allerdings ist das Element `<guide>` in EPUB Version 2 nur unter Vorbehalt zu empfehlen.

⁴⁹ Vgl. Data2type (o.J.): ePUB. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

Tabelle 1. Liste der Referenz-Typen innerhalb von <guide>⁵⁰

Wert	Beschreibung
cover	Buchumschlag / Buchcover
title-page	Titelseite
toc	Inhaltsverzeichnis (table of contents)
index	Stichwortverzeichnis
glossary	Glossar
acknowledgements	Danksagung
bibliography	Literaturverzeichnis
colophon	Kolophon
copyright-page	Impressum-Seite
dedication	Widmung
epigraph	Epigraph
foreword	Geleitwort
loi	Abbildungsverzeichnis
lot	Tabellenverzeichnis
notes	Anmerkungen
preface	Vorwort
text	Erste echte Inhaltsseite

In EPUB Version 3 sind die Elemente <guide> und <tour> als veraltet gekennzeichnet, dass sie nicht mehr verwendet werden sollen. Dafür kann optional das neue Element <bindings> folgen. Das Element <bindings> definiert eine Reihe von benutzerdefinierten Routinen (handler) für Medientypen, die von dieser Spezifikation standardmäßig nicht unterstützt sind.

Daraufhin darf in Version 3 noch Weiteres hinzugefügt werden. Zum Beispiel, das Unterelement <meta> erweitert die Metadaten der Publikation, indem es zusätzliche

⁵⁰ Vgl. Data2type (o.J.): ePUB. Online im Internet. <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

detaillierte Ebene(n) anbietet. Das Element <link> wird verwendet, um Ressourcen mit einer Publikation, wie Metadatensätze zu verknüpfen.⁵¹

3.5.4.2 Inhaltsverzeichnis

Die EPUB-Standard Version 2 hat den vom DAISY-Consortium publizierten Standard, NCX DTD, eingeführt.⁵² Die Abkürzung NCX steht für "**N**avigation **C**ontrol file for **X**ML **a**pplications". NCX ist eigentlich eine XML-Datei. Wie bereits erwähnt, geht es bei dieser Datei um die Navigation. NCX ermöglicht in komplexen Büchern eine hierarchische Gliederung darzustellen und eine seitenorientierte Navigation abzubilden. Im Element <ncx> werden für das ganze Dokument der Default-Namensraum xmlns="http://www.daisy.org/z3986/2005/ncx/" deklariert und noch weitere drei Attribute angegeben: Die Version, die Sprache und die Leserichtung. NCX DTD verlangt vier Bezeichner in <head>-NCX: „dtb:uid“ ist die eindeutige ID, die mit dem Wert des Attributs *unique-identifier* von <package> übereinstimmen sollte; "dtb:depth" spiegelt die hierarchische Tiefe der NCX-Datei wieder; "dtb:totalPageCount" und "dtb:maxPageNumber" gelten nur für Papierbücher und für elektronische Dateien bleiben die Werte von beiden „0“. Der Titel wird im Element <docTitle> mithilfe des Unterelementes <text> angegeben. <navMap> als der wesentliche Teil der NCX-Datei beschreibt die eigentliche Hierarchie der Navigation. Jedes seiner Unterelemente <navPoint> stellt den Titel mit <navLabel> und die Inhaltsreferenz mit <content> auf. Eine Verschachtelung im Element <navPoint> ist erlaubt. Außerdem besitzt das Element zwei Pflichtattribute. Das Attribut "playOrder", das einen ganzzahligen Wert der Reihenfolge von <itemref> enthalten muss, und das Attribut "id", welches mit dem Wert des Attributs *dc:identifier* aus der OPF-Datei übereinstimmen sollte. Darunter gibt das <navLabel>-Element den jeweiligen Titel an. Des Weiteren verweist das Pflichtattribut *src* innerhalb des <content>-Elementes auf die eigentliche Inhaltsdatei, die in OPF-<manifest> deklariert ist oder sich mit einem Inhaltsfragment zeigen.

⁵¹ Vgl. Idpf (2011): EPUB Publications 3.0. Online im Internet: <http://www.idpf.org/epub/30/spec/epub30-publications.html> [09.09.2015]

⁵² Vgl. DAISY (2015): DAISY. Online im Internet: <http://www.daisy.org> [09.09.2015]

Neben der logischen-hierarchischen Gliederung ist in der aktuellen EPUB-Version auch eine seitenorientierte Gliederung möglich. Das Element `<div>` mit den Attributen `class` und `id` wird dafür angewandt, die Seitenzahl der Inhaltsdatei anzugeben. Demgemäß werden die Seitenmetadaten in der NCX-Datei geändert und das Element `<pagelist>` befüllt. NCX stammt aus einem anderen Standard. Demzufolge gibt es Wiederholungen der gleichen Informationen in NCX und OPF. Beim Lesen wird der Wert an verschiedenen Stellen aufgewiesen. Ein Fehler kann durch die manuelle Kodierung auftreten. Deshalb muss man beim Erstellen die Konsistenz behalten. Eine NCX-Datei kann wie im Folgenden gezeigt, formuliert werden (Siehe Quellcode 5).

Quellcode 5. NCX-Datei-Inhalte-Beispiel⁵³

```
<?xml version="1.0" encoding="UTF-8"?>
<ncx xmlns="http://www.daisy.org/z3986/2005/ncx/" version="2005-1" xml:lang="de"
dir="ltr">
  <head>
    <!--Die folgenden vier Metadatenelementen sind für alle NCX-Dokumente
erforderlich, einschließlich diejenigen entspannten Einschränkungen von OPS 2.0-->
    <meta name="dtb:uid" content="OPF-BuchId"/>
    <meta name="dtb:depth" content="Anzahl Ebenen"/>
    <meta name="dtb:totalPageCount" content="0"/>
    <meta name="dtb:maxPageNumber" content="0"/>
  </head>
  <docTitle>
    <text>Buchtitel</text>
  </docTitle>
  <docAuthor>
    <text>Vorname Nachname</text>
  </docAuthor>
  <navMap>
    <navPoint id="Eintrag-ID" playOrder="1">
      <navLabel>
        <text>Eintragtext</text>
```

⁵³ Vgl. Data2type (o.J.): ePUB. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

```

        </navLabel>
        <content src="Verweis zu Inhaltsdatei"/>
    </navPoint>
    ...
</navMap>
</ncx>

```

Der Inhalt von gedruckten klassischen Büchern wird mit einer passenden Menge auf einzelne Seite aufgeteilt und strukturiert. Bei E-Books von zurückfließbaren Inhalten ist das sinnlos. Die Seitenangaben sind je nach Plattform und stilistischen Einstellungen, wie beispielsweise Schriftgröße und Seitenrand, völlig unterschiedlich. Angemessene Verweise auf beliebige Stellen sind deshalb mit herkömmlichen Methoden nicht erstellbar. Die Inhalte sind normalerweise durch einzelne Inhaltsdateien sinnvoller kapitelweise aufgeteilt. Von daher können die Verweise anhand der Strukturen in E-Books, wie Kapitel, Abschnitt, Absatz usw., praktisch erfolgen.

Die Navigation von EPUB 3 ist ähnlich wie in EPUB 2 strukturiert, wird aber davon getrennt behandelt. Für EPUB 3 ist die NCX-Datei optional. EPUB 3 definiert eine neue menschenlesbare Grammatik für Navigationsinformationen innerhalb einer Publikation als ein normales XHTML-Inhaltsdokument. Die für die Navigation wesentlichen Bestandteile sind durch `<nav>` herausgeführt. In dem Navigationsdokument entspricht das Element `toc-nav` dem Element `navMap` in der NCX-Datei. Jedes `nav`-Element kann eine optionale Überschrift enthalten, die den Titel der Navigationsliste andeutet. Jedes Listenelement `` der geordneten Liste `` vertritt eine primäre Überschrift. Die Reihenfolge aller Inhaltsdateien wird durch Listenelemente festgelegt. Jedes Listenelement muss entweder ein Unterelement `<a>` oder `` enthalten. Das Attribut `a` referenziert eine einzelne Datei im Verzeichnis, während `span` dazu dient, Text aufzunehmen, mit dem die Liste durch Überschriften gruppiert oder strukturiert werden kann. Das Element `<a>` kann gegebenenfalls von einem `` gefolgt werden. Daraufhin entspricht die EPUB-Eigenschaft `landmarks-`

nav dem Element <guide> in der NCX-Datei.⁵⁴ Das sieht zum Beispiel wie folgt aus (Siehe Quellcode 6):

Quellcode 6. Beispiel XHTML-Inhaltsverzeichnis-Datei⁵⁵

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
    "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
    xmlns:ops="http://www.idpf.org/2007/ops" xml:lang="de">
  <head>
    <title>Buchtitel</title>
  </head>
  <body>
    <h1>Eintragstext</h1>
    <nav epub:type="toc" id="toc">
      <ol>
        <li>
          <a href="Verweis zu Inhaltsdatei">Inhalt1</a>
        </li>
        <ol>
          <li>
            <a href=" Verweis zu Grafikdatei">Grafik1</a>
          </li>
        </ol>
        <li>
          <a href=" Verweis zu Inhaltsdatei">Inhalt2</a>
        </li>
        ...
      </ol>
    </nav>
  </body>
</html>
```

⁵⁴ Vgl. Idpf (2011): EPUB Publications 3.0. Online im Internet: <http://www.idpf.org/epub/30/spec/epub30-publications.html> [09.09.2015]

⁵⁵ Vgl. Data2type (o.J.): ePUB. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

3.5.4.3 Stilvorlagen

Die Anforderung der Stilvorlagen von EPUB sind durch CSS (Abkürzung von Englisch: Cascading Style Sheet) erfolgt. Die Grundlinie vom CSS-Profil in EPUB Version 3 ist CSS 2.1. Mit einigen Ausnahmen umfasst das CSS-Profil fast alle Konzepte in CSS 2.1.⁵⁶ Als eine Formatierungssprache für HTML- und XML-Dokumente hat CSS eine ganz andere Struktur und bietet damit vielfältige Darstellungsmöglichkeiten.

Die Eigenschaften in CSS sind von Regelsätzen organisiert. Ein Regelsatz besteht aus einem Selektor oder einer Gruppe von Selektoren und einem durch geschweifte Klammern begrenzten Bereich, in dem beliebige durch Semikolon getrennte Deklarationen aufgeführt sind. Jeder Deklaration werden eine Eigenschaft und ein Wert zugewiesen. CSS weist Selektoren für Elemente und Attribute von XML- bzw. XHTML-Dokumente auf. Im Folgenden ist ein Beispiel eines Regelsatzes aus einer Vorlagendatei:

```
Selektor Eins, Selektor Zwei { Eigenschaft: Wert; Eigenschaft2: Wert; }57
```

3.6 Software für die Erstellung von EPUB-Dokumenten

Ein komprimiertes Archiv-Dokument ist nicht direkt mit einem Texteditor zu bearbeiten. Es gibt verschiedene Arten von Programmen und Tools auf dem Markt, mit denen ein EPUB-Dokument aus anderen Formaten erstellt werden kann. Mit einigen Programmen kann ein EPUB-Dokument Schritt für Schritt erstellt werden. Dabei wird nicht viel Fachwissen der Informatik vorausgesetzt. „Es zeichnet sich aber schon ab, dass die einschlägigen Hersteller von Publikations- und Layout-Software, allen voran Adobe, in ihre Standardprogramme alle Funktionen für die Erstellung valider EPUB-Dateien integrieren“ [Handner 2011].

⁵⁶ Bei der Stilvorlage werden einige Eigenschaften von CSS 2.1 ausgenommen und wenige ergänzt. Ausführliche Informationen siehe IDPF.

⁵⁷ Vgl. W3S (o.J.): W3S. CSS Tutorial. Online im Internet: <http://www.w3schools.com/css/> [09.09.2015]

Adobe InDesign ist ein professionelles Layout- und Satzprogramm für das DPT (Desktop Publishing), mit dem man ein Dokument gut aufbereiten kann. Adobe InDesign ist der Nachfolger von Adobe PageMaker, der für das professionelle Drucken entwickelt wurde. Seit der Version 5 von Adobe PageMaker wird EPUB unterstützt.⁵⁸ Der Ersteller kann von Anfang an selbständig mit Adobe InDesign sein eigenes Layout, Deckblatt, Icon, Steuerelement konstruieren. Das Programm bietet dem Benutzer viele Möglichkeiten für das DPT.

Calibre ist eine freie Software-Suite zur Verarbeitung und Verwaltung von E-Book-Daten. Calibre kann die meisten derzeit relevanten Formate konfiguriert darstellen, ausdrucken und konvertieren. Die Inhalte lassen sich beim Erstellen aufbereiten.⁵⁹

Sigil ist ein Open-Source-Editor. Dieser unterstützt sowohl WYSIWYG als auch das direkte Editieren des HTML-Quellcodes. Ziel des Projekts ist es, einen benutzerfreundlichen Editor für EPUB-Dokumente zu erschaffen. Sigil kann automatisch *toc* generieren, Bilder einfügen und Metadaten bearbeiten.⁶⁰ Ein EPUB-Dokument wird in Sigil durch das Öffnen und Speichern in seiner eigenen Struktur geändert. (Siehe Abbildung 3).

⁵⁸ Vgl. Adobe InDesign (2015): Adobe InDesign. Online im Internet: https://de.wikipedia.org/wiki/Adobe_InDesign [09.09.2015]

⁵⁹ Vgl. Calibre (2015): Calibre. Online im Internet: <https://de.wikipedia.org/wiki/Calibre> [09.09.2015]

⁶⁰ Vgl. Sigil. Online im Internet: <https://de.wikipedia.org/wiki/Sigil> [09.09.2015]

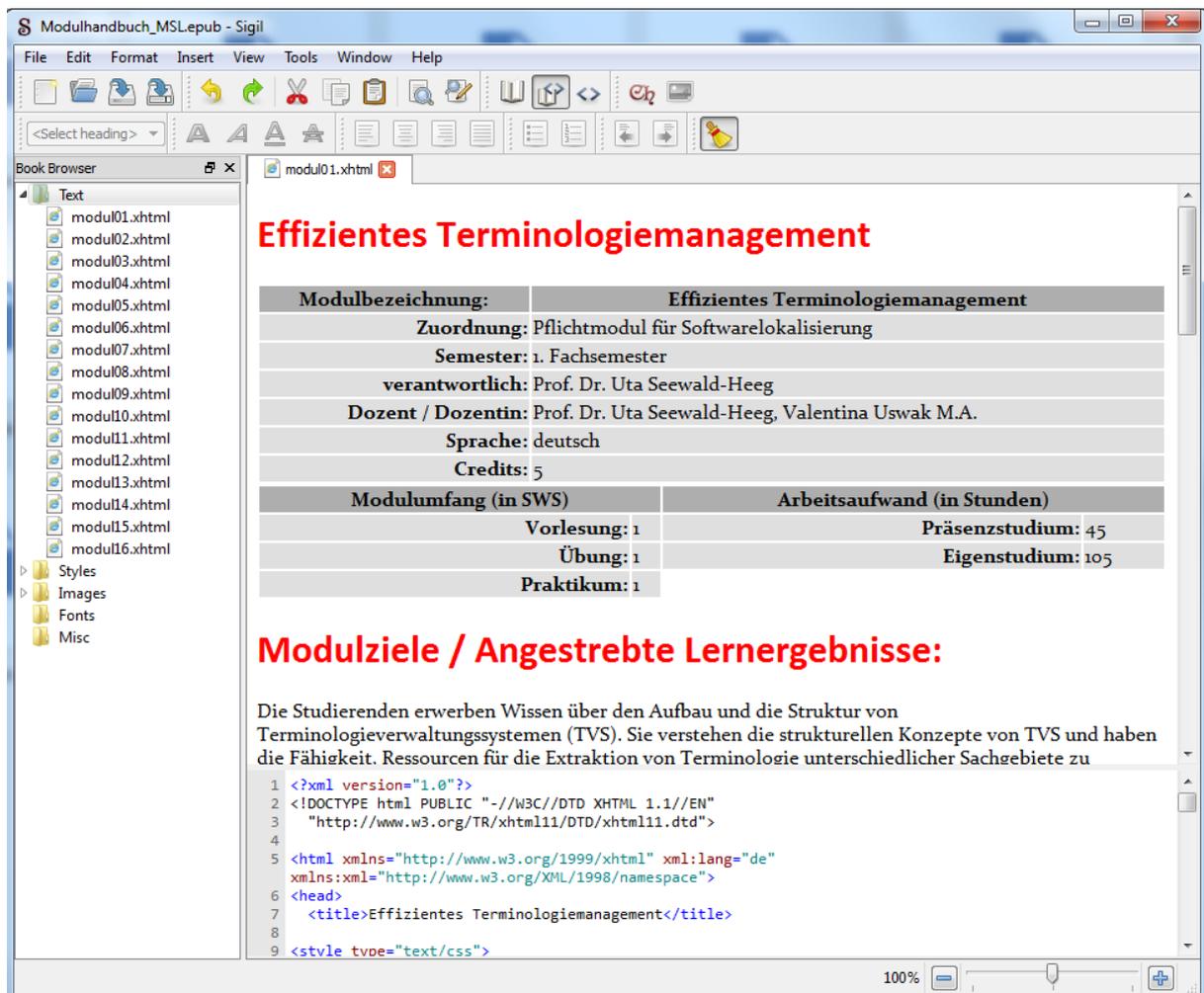


Abbildung 3. Erstellen eines EPUB-Dokuments mithilfe von Sigil

EpubSTAR, EPUB-Builder und EPUBMaker sind E-Book-Editoren, die aus China stammen. Daher werden diverse chinesische Schriftarten von diesen Programmen gut unterstützt. Allerdings sind die Leistungen von Schriftsätzen und der Analyse von Stilvorlagen noch unbefriedigend.⁶¹

Die Handhabung aller Programme ist ähnlich. Zuerst soll der Ressourcen-Ordner bereitgestellt werden. Darin sind die Buchinhalte am besten kapitelweise und eventuelle Verweise im HTML-Format fertiggestellt. Der Ordner und die darin enthaltene Dateien kann man entsprechend des Buches benennen. Mit Hilfe der ausgewählten Software für das Erstellen von EPUB legt man zuerst ein neues

⁶¹ Vgl. Niu (2012): EPubSTAR. Online im Internet im Chinesischen: http://www.by-smart.com/eread/eread_11.html [09.09.2015]

EPUB-Dokument an. Im Dialogfeld des Editors kann man die Metadaten vom Buchnamen bis hin zu den Beschreibungen eingeben. Danach importiert man die vorhandenen Dateien aus dem Ressourcen-Ordner in das Programm. Das Programm wird automatisch eine Dateiliste (manifest, spine usw.) generieren. Dann kann man die Informationen des Inhaltsverzeichnisses aufnehmen. Durch „Alles Speichern“ kann man schließlich das EPUB-Dokument erhalten.

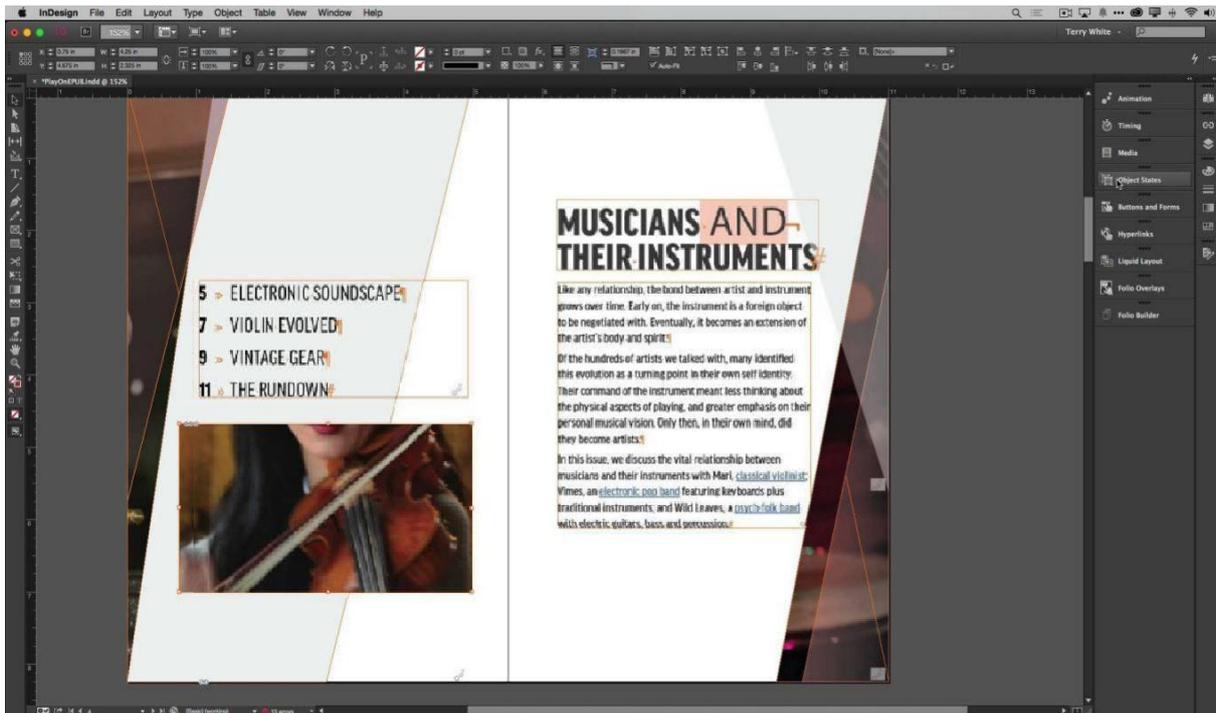


Abbildung 4. Erstellen eines EPUB-Dokuments mithilfe von Adobe InDesign⁶²

⁶² Vgl. T. White (2014): How To Create a Slideshow in InDesign CC for your Interactive ePUB. Screenshot aus Youtube. Online im Internet: <https://www.youtube.com/watch?v=Poj5EYwxcMk> [09.09.2015]

4. Erstellung eines EPUB-Dokuments am Beispiel der Modulbeschreibungen für den Masterstudiengang Softwarelokalisierung

4.1 Internationalisierungsanforderungen

Wegen des Vorantreibens der interkulturellen Kommunikation und Kooperation in den akademischen und wirtschaftlichen Bereichen, stehen alle Firmen der Herausforderung gegenüber, ihre Produkte und Dienstleistungen anderen Zielmärkten anzubieten. Dabei steht die Internationalisierung im Mittelpunkt der Entwicklung eines Produktes.

Heute sind die Lokalisierung (Front-End Anpassung) und die Internationalisierung (Back-End Technologien) zwei sich ergänzende Prozesse der Entwicklung von kundenspezifischer Globalisierung. In der Industrieentwicklung ist die Lokalisierung nicht isoliert, sondern ein Bestandteil der voneinander abhängenden Prozesse, die gemeinsam GILT genannt werden. Das Akronym GILT kommt aus dem Englischen und steht für „Globalization“, „Internationalization“, „Localization“ und „Translation“. Die Globalisierung (kurz „G11n“) konzentriert sich in erster Linie auf die organisatorische Ebene, die sich auf alle zu begegneten geschäftlichen Entscheidungen und Tätigkeiten bezieht. Die gesamten Vorgänge der Entwicklung, Übersetzung und Vermarktung sind in der Globalisierung enthalten. Der Zweck der Globalisierung ist es, alle Kunden weltweit zu unterstützen und alle Prozesse der Produktion, der Ausbreitung und des Gebrauchs reibungslos zu gewährleisten. Internationalisierung (I18n) ist der Vorgang, der ein Produkt einfacher für ein Zielland bzw. ein Zielmarkt erstellen lässt. Dafür werden die kultur- und landspezifischen Elemente von dessen Kern entfernt und versucht, das Produkt so neutral wie möglich zu halten. Idealerweise passiert die Internationalisierung in der Entwicklungsphase, mit anderen Worten, die Internationalisierung sollte immer der Lokalisierung vorausgehen.⁶³ Produkte und ihre Dokumentationen werden durch die

⁶³ Vgl. N. Singh (2005): A. Pereira. The culturally customized web site. Customizing web sites for the global marketplace. 1. Auflage. Elsevier Butterworth-Heinemann.

Internationalisierung möglichst unabhängig von irgendeiner Kultur oder Sprache entworfen. In diesem Schritt werden also die Voraussetzungen für eine gute Lokalisierung gewährleistet. Die Programmierung des Programmskeletts (Skeleton)⁶⁴ braucht bei einem internationalisierten Programm bei der Lokalisierung nicht neu erstellt oder geändert zu werden.

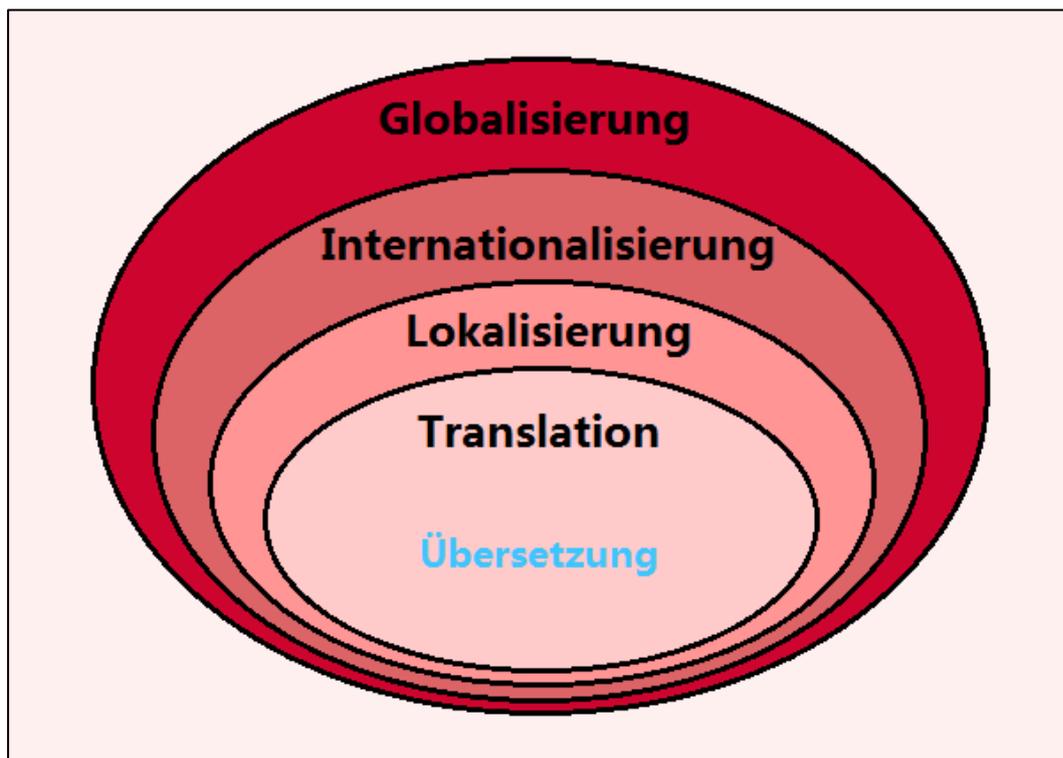


Abbildung 5. GILT⁶⁵

Bei der Erstellung des EPUB-Dokuments muss man auf die Kulturneutralität achten. Neben dem Fließtext sind die Kodierung von XHTML und XML universal standardisiert. Damit das Dokument später einfach in andere Sprachen übertragen werden kann, sollten wenige kulturspezifische Elemente in den Ausgangsmaterialien übernommen werden. Weil die Ausgangsmaterialien in einer Sprache verfasst wurden, ist es unmöglich, Produkte ohne etwas kulturelle Eigenschaften darzustellen. Das heißt, dass der Text im Deutschen die grundlegenden Regeln der Sprache befolgen muss. Zum Beispiel, anders als die arabische Textrichtung, ist Deutsch eine

⁶⁴ Skeleton-Datei ist der nicht zu lokalisierende bzw. nicht zu übersetzende Teil eines Softwareproduktes.

⁶⁵ Vgl. U. Seewald-Heeg (2009): Lokalisierung Grundlagen. Folien für die Lehrveranstaltung Lokalisierung Grundlagen an der Hochschule Anhalt.

der von links nach rechts laufenden Schriften. Für die Lokalisierungsprozesse ist die gute Bereitschaft der Kodierung von Quelldateien bedeutsam. Das beeinflusst die Extrahierung und Segmentierung der Ausgangstexte und gleicherweise den Export der Zieldatei.

Die Inhalte eines EPUB-Dokuments sind XHTML-Dateien, während deren Beschreibungen XML-Dateien sind. Das Erstellen von EPUB-Dokumenten ist nicht kompliziert, sodass es jeder machen kann und keine speziellen Hilfsmittel dafür notwendig sind. Denn die Inhalte basieren auf (X)HTML, kann sie ebenfalls jeder lesen.

4.2 Verwendete Software-Produkte

Da es keine perfekte Software gibt, die die unterstützten Dokumenten ohne Probleme unmittelbar in EPUB konvertieren kann. Um ein sauberes und gut strukturiertes EPUB-Dokument herzustellen, wurden die speziellen Werkzeuge zum EPUB-Erstellen in diesem Projekt verzichtet, sondern direkt die Quellcodes geschrieben.

Zum praktischen Arbeiten wird mindestens eine Texteditor-Software benötigt, die die Bearbeitung und Überprüfung von XHTML- und XML-Dateien unterstützt. *Notepad ++* und *Altova XML Spy* wurden beim Erstellen von Inhalt- und Darstellungsdateien des Modulhandbuchs verwendet. Mit *Altova XML Spy* wird überprüft, ob die erzeugten Dateien alle Regeln des XHTML- und der XML-Standards einhalten. Außerdem ist ein Packprogramm erforderlich, mit dem ein Archiv im ZIP-Format erstellt werden kann. Zudem können damit verborgene Informationen zugänglich gemacht werden. Das Packprogramm *WinRAR* hat beim Generieren der ZIP-Datei geholfen. Bei Bedarf kann auch ein Grafikeditor verwendet werden, um grafische Bestandteile zu bearbeiten.

4.3 Vorbereitung der Inhaltsdateien und Funktionsdateien

4.3.1 Konvertierung und Vorbereitung von Inhaltsdateien

Vor dem Erstellen eines EPUB-Dokuments müssen zunächst alle Dateien vorbereitet werden. Die Inhaltsdateien eines EPUB-Dokuments sind im Allgemeinen XHTML-Dateien, die die meisten zu lokalisierenden Texte enthalten. Die Ausgangsmaterialien liegen als Word-Dateien vor. Am Anfang wurde versucht den Aufbau der XHTML-Dateien zusammen mit den Inhalten zu skizzieren. Dafür wurden alle Details aus den Word-Dateien in XHTML nachgebildet. Dabei halfen die Funktionen von Microsoft Word, Absatzmarken und sonstige ausgeblendete Formatierungssymbole anzuzeigen und alle Formatierungen hervorgehoben zu erkennen.

Microsoft Word bietet mit eingebauten Funktionen die Möglichkeit, eine Word-Datei in das XHTML-Format zu konvertieren, jedoch aus Gründen der besseren Qualität wurden die XHTML-Dateien manuell konzipiert (Siehe Abbildung 6).

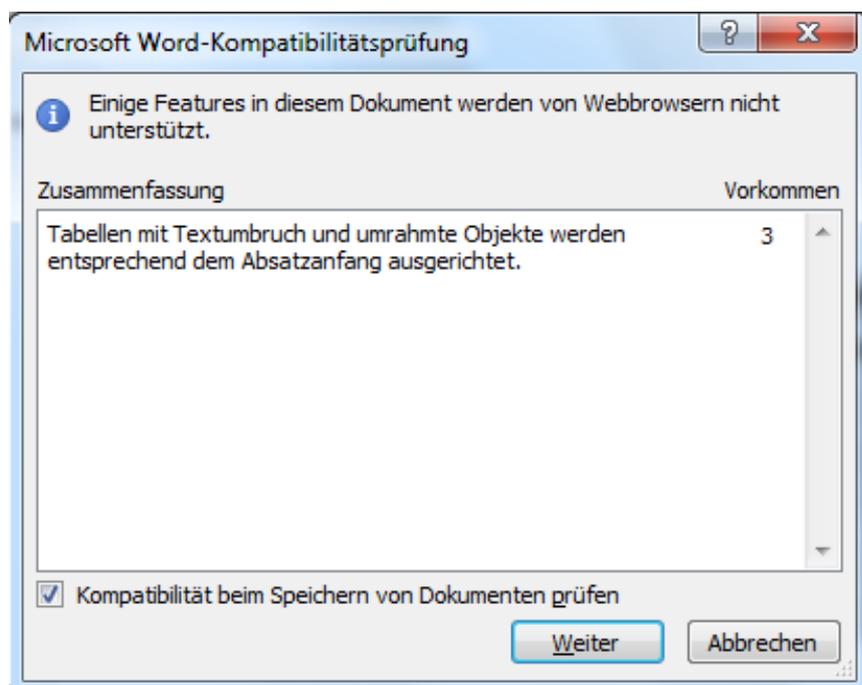


Abbildung 6. Warnmeldung bei der Konvertierung von Worddateien in das PDF-Format

Neben der XML-Präambel in der ersten Zeile, dem Dateityp-Element und dem Namensraum am Anfang der Datei ist noch die Sprache in der XHTML-Inhaltsdatei zu definieren. Das Attribut *xml:lang* legt die Sprache der Datei fest. Die Lokalisierungs- und Terminologiewerkzeuge werden durch das Element mit dem Attribut "de" für die Sprache Deutsch eingestellt. (Siehe Quellcode 7)

Quellcode 7. Der Kopf von XHTML-Inhaltsdateien⁶⁶

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
    "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="de">
```

Einige Zeichen werden in der Auszeichnungssprache reserviert und können deshalb nicht direkt im Quellcode verwendet werden, besonders am Anfang eines Strings. Die reservierten Zeichen in der Auszeichnungssprache müssen mit Zeichen-Entitäten ersetzt werden. Wenn die reservierten Zeichen im Fließtext (zwischen dem Anfangs- und dem End-Tag eines Elements) verwendet werden, könnte der Browser die Zeichen mit Tags u.ä. verwirren und gar nicht lesen. Allerdings kann nicht immer das gewünschte Ergebnis ausgeliefert werden, da dieses von vielen Faktoren abhängig ist. Beispielsweise vom verwendeten Web-Browser oder von der Verbindung zur Datenbank. Um Dateien in guten Zustand für Lokalisierung und zur Darstellung auf unterschiedlichen Geräten vorzubereiten, sollen die Sonderzeichen vor allem in HTML-Code (Unicode⁶⁷) umgeschrieben werden (siehe Tabelle 2), sonst sind diese Zeichen eventuell in manche Web-Browsern nicht darstellbar oder von manchen Anwendungen nicht zugelassen (siehe Abbildung 7). Oft werden Umlaute von Web-Browsern falsch angezeigt. Weil die Texte in XHTML aus Word-Dateien kopiert werden und manche Tags, Style-Attribute u.a. von XHTML Sonderzeichen enthalten, wie Kaufmanns-Und-Zeichen, eckige Klammern, Anführungszeichen usw., ist ein globaler Ersatz von Zeichen in der XHTML-Datei ohne Störung nicht möglich.

⁶⁶ Vgl. Data2type (o.J.): ePUB. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

⁶⁷ Mit Unicode werden so ziemlich alle Zeichencodierungsprobleme dieser Welt gelöst. Und weil Webseiten potentiell mit allen Schriften dieser Welt genutzt werden, ist für HTML 4.0 und XML (und damit auch XHTML) festgelegt, dass grundsätzlich alle in Unicode definierten Zeichen vorkommen dürfen. (Vgl. selfhtml. Zeichencodierung. Online im Internet)

Daher sollten die Sonderzeichen vor dem Import in eine Word-Datei durch HTML-Entitäten ersetzt werden.

Tabelle 2. Im Projekt auftretende Codes für HTML-Sonderzeichen⁶⁸

Zeichen	Ä	Ä	ö	Ö	ü	Ü	-
HTML-Entity	ä	Ä	ö	Ö	ü	Ü	−
Zeichen	ß	<	>	“	”	„	-
HTML-Entity	ß	<	>	“	”	„	–
Zeichen	€	&	é	ç	®	'	—
HTML-Entity	€	&	é	Ç	®	’	—

This page contains the following errors:

```
error on line 119 at column 11: AttValue: " or ' expected
error on line 119 at column 11: Encoding error
```

Below is a rendering of the page up to the first error.

Abbildung 7. Fehlermeldung bei Sonderzeichen in XHTML

Es wurden verschiedene waagerechte Striche in unterschiedlichen Formen in den Materialien angewandt. In Deutschland werden Minuszeichen häufiger als Viertelgeviertstriche (als Bindestriche in Deutschland verwendet) verwendet. Die beiden waagerechten Striche sehen normalerweise gleich aus. Der Unterschied ist nur bei bestimmten Schriftarten und Programmen erkennbar. Abbildung 8 zeigt die verschiedenen Darstellungen in einer Webseite. Das Minuszeichen ist etwas länger

⁶⁸ Vgl. selfhtml. Zeichencodierung. <<http://wiki.selfhtml.org/wiki/Zeichencodierung#ASCII>>

als der Viertelgeviertstrich und steht ein bisschen höher als der Viertelgeviertstrich.⁶⁹ In einer Word-Datei kann man nicht einfach ein Minuszeichen von einem Viertelgeviertstrich unterscheiden. Für den Halbgeviertstrich (Gedankenstrich) gibt es keine spezielle Taste auf der deutschen Tastatur. Mit dem Kürzel „Strg“ + „-“⁷⁰ ist der Gedankenstrich aber auch mit der deutschen Standardtastatur verfügbar. Alle Gedankenstriche werden als „–“ in HTML ersetzt.

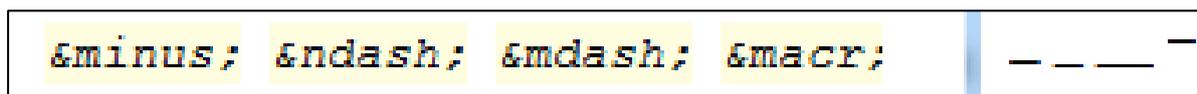


Abbildung 8. Unterschied zwischen Minus-, en/em Gedanken⁷¹- und Überstrich

4.3.2 Ermittlung von Inkonsistenten

In der Vorbereitungsphase werden Schreibfehler und Inkonsistenzen herausgefunden. Zum Beispiel werden die Angaben für den Arbeitsaufwand manchmal mit der Einheit "h" für Stunde angegeben und manchmal nicht (siehe Abbildung 9). Der Fehler wurde dadurch beseitigt, dass alle Angaben in den Tabellen „Arbeitsaufwand“ ohne Einheiten dargestellt wurden.

⁶⁹ Vgl. T. Otte(2007): Bindestrich in HTML. Online im Internet: <http://tobias-otte.de/2007/05/bindestrich-in-html/> [09.09.2015]

⁷⁰ Das Zeichen „-“ ist das Minuszeichentaste auf dem numerischen Tastenblock rechts auf der Tastatur.

⁷¹ „en/em Gedankenstrich“ sind Gedankenstrich Breite n und m. Die Maßeinheiten en und em werden im Schriftsatz zur Bestimmung der Zeichenbreite und Höhe in Abhängigkeit von der Schriftgröße verwendet [em 2014]. Der Bindestrich entspricht „hyphen“ im Englischen. Der „en dash“ (ungefähr die Breite des Großbuchstabens N) „entspricht in etwa dem deutschen Gedankenstrich“ [sprachenquilt 2009]. Der „em dash“ (Deutsch: Geviertstrich, ungefähr die Breite des Großbuchstabens M) wird allerdings im Deutschen in der Regel nicht verwendet.

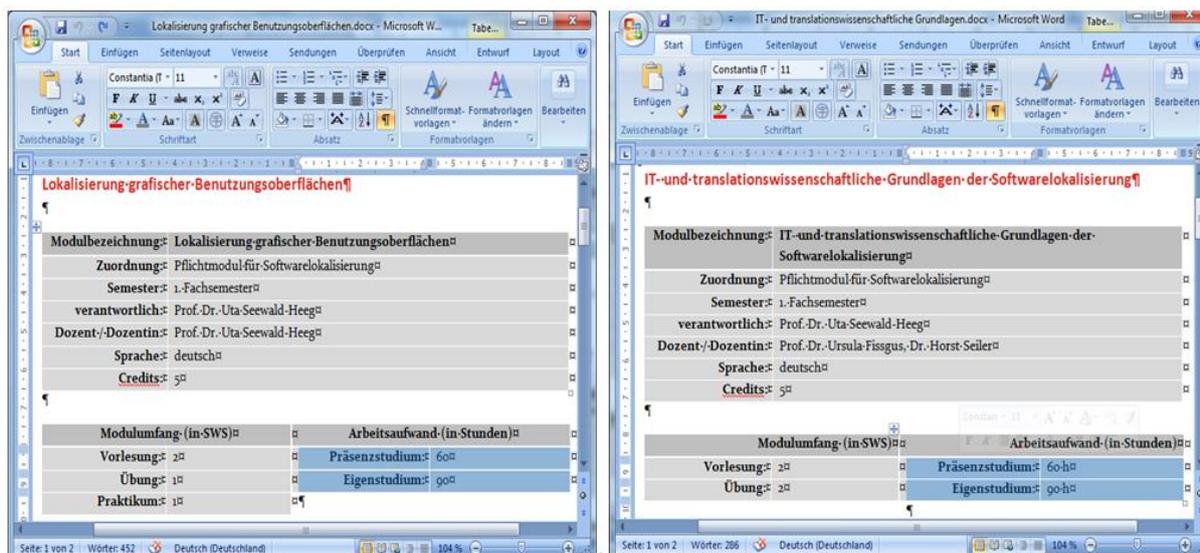


Abbildung 9. Inkonsistenz der Ausgangsmaterialien

Die Style-Informationen werden im Laufe des Erstellens der Inhaltsdateien kurz und knapp für alle Dateien eingebettet. Daher ist die Datei *stylesheet.css* nicht mehr notwendig. Der Abschnitt `<style>` mit den darin festgestellten Werten kann bei Bedarf als Stilvorlage für mehrere Dateien verwendet werden. Die für die ganze E-Book-Datei geltenden CSS-Regeln werden innerhalb des `<head>`-Elements formuliert. Die grundlegende Schriftart des Dokuments wird als *Constantia* und die Titelschrift als *Calibri* wie in den Ausgangsmaterialien eingestellt. Um die Tabelle in einer geeigneten Darstellungsästhetik auf unterschiedlichen Geräten zu zeigen und das Druckformat anzupassen, wird die maximale Breite von Tabellen mit 600 Pixel eingerichtet. Im digitalen Medium ist es nicht vorhersehbar, in welcher Schriftgröße ein Dokument angezeigt wird. Daher ist es empfehlenswert, Schriftgröße und Abstand zwischen einzelnen Textzeilen eines Absatzes mit Prozentzahlen relativ festzulegen. Die übrigen Angaben haben automatisch die Default-Formate der jeweiligen Elemente von XHTML beibehalten.

Als Beispiel der Inhaltsstruktur wurde der Quellcode vom Modul Effizientes Terminologiemanagement angehängt (Siehe Quellcode 8), für die anderen 15 Dokumente sind nur die in der Skeleton-Struktur eingebetteten Inhalte unterschiedlich:

Quellcode 8. Inhaltsstruktur der Webseite des Moduls Effizientes Terminologiemanagement⁷²

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
    "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="de">
  <head>
    <title> Effizientes Terminologiemanagement </title>
    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8"/>
    <style type="text/css" xml:space="preserve">
      table { max-width: 600px; width: 100%}
      body{font-family:Constantia}
      h1 {font-family:Calibri}
      li {line-height: 150%}
    </style>
  </head>
  <body>
    <h1 style="color:#FF0000">Effizientes Terminologiemanagement</h1>
    <table>
      <tbody>
        <tr style="background-color:#AEAEAE">
          <th>Modulbezeichnung:</th>
          <th>Effizientes Terminologiemanagement </th>
        </tr>
        <tr style="background-color:#DFDFDF">
          <td align="right">
            <b>Zuordnung: </b>
          </td>
          <td>Pflichtmodul f&uuml;r
Softwarelokalisierung</td>
        </tr>
        <tr style="background-color:#DFDFDF">
          <td align="right">
            <b>Semester: </b>
          </td>
          <td>1. Fachsemester </td>
        </tr>
      </tbody>
    </table>
  </body>
</html>

```

⁷² Vgl. Data2type (o.J.): ePUB. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

```

        <tr style="background-color:#DFDFDF">
            <td align="right">
                <b>verantwortlich: </b>
            </td>
            <td> Prof. Dr. Uta Seewald-Heeg </td>
        </tr>
        <tr style="background-color:#DFDFDF">
            <td align="right">
                <b>Dozent / Dozentin: </b>
            </td>
            <td> Prof. Dr. Uta Seewald-Heeg, Valentina Uswak
M.A.</td>
        </tr>
        <tr style="background-color:#DFDFDF">
            <td align="right">
                <b>Sprache: </b>
            </td>
            <td>deutsch</td>
        </tr>
        <tr style="background-color:#DFDFDF">
            <td align="right">
                <b>Credits:</b>
            </td>
            <td>5 </td>
        </tr>
    </tbody>
</table>
<p/>
<table>
    <tbody>
        <tr style="background-color:#AEAEAE">
            <th colspan="2" align="center"> Modulumfang (in
SWS)</th>
            <th colspan="2" align="center"> Arbeitsaufwand (in
Stunden)</th>
        </tr>
        <tr style="background-color:#DFDFDF">
            <td align="right">
                <b>Vorlesung: </b>

```

```

        </td>
        <td>1</td>
        <td align="right">
            <b>Pr&auml;senzstudium: </b>
        </td>
        <td>45</td>
    </tr>
    <tr style="background-color:#DFDFDF">
        <td align="right">
            <b>&Uuml;bung: </b>
        </td>
        <td>1 </td>
        <td align="right">
            <b>Eigenstudium:</b>
        </td>
        <td>105 </td>
    </tr>
    <tr>
        <td align="right" style="background-
color:#DFDFDF">
            <b>Praktikum:</b>
        </td>
        <td style="background-color:#DFDFDF">1</td>
        <td align="right">
            <b/>
        </td>
        <td/>
    </tr>
</tbody>
</table>
<p/>
<h1 style="color:#FF0000">Modulziele / Angestrebte
Lernergebnisse:</h1>
    <p> Die Studierenden erwerben Wissen &uuml;ber den Aufbau und die
Struktur von Terminologieverwaltungssystemen (TVS). Sie verstehen die
strukturellen Konzepte von TVS und haben die F&auml;higkeit, Ressourcen f&uuml;r
die Extraktion von Terminologie unterschiedlicher Sachgebiete zu ermitteln und
qualitativ zu evaluieren. Die Studierenden lernen terminologische Normen und
Formate zum Austausch terminologischer Daten. Die Studierenden kennen Konzepte zum

```

Aufbau von Terminologiebeständen in Unternehmen, sie können Importskripte erstellen und Algorithmen zur Dublettenbereinigung entwickeln.

Die Studierenden können auf der Grundlage unternehmensspezifischer Voraussetzungen Workflow-Szenarien entwickeln und ein unternehmensweites Terminologiemanagement konzeptionell entwickeln.

Inhalt:

- Terminologieverwaltungssysteme (TVS)
- Integration von TVS in Anwendungen zur Übersetzung und Lokalisierung
- Nutzung von Terminologie in der Qualitätssicherung
- Terminologische Vorgaben in Stilrichtlinien (Style Guides)
- Aufbereitung terminologischer Bestände für das maschinelle Lektorat
- Austausch terminologischer Daten
- Problematisierung von Datenhaltung und Kodierung der terminologischen Daten.

 Terminologische Normen

Voraussetzungen:

- keine

Studien-/Prüfungsleistungen, Prüfungsformen:

Die Studienleistung wird anhand einer mündlichen Prüfung vom 25 Minuten Dauer bewertet.

Eingesetzte Medienformen:

- Online-Materialien im Lernmanagementsystem moodle
- Präsentationsunterlagen (Powerpoint)
- Terminologieverwaltungssysteme, online und offline

Literatur:


```

        <li> Arntz, Reiner/Picht, Heribert/Mayer, Felix :
Einführung in die Terminologearbeit. lms, Hildesheim, Zürich,
2002.</li>

        <li> Cabrera, M. Teresa: Terminology. Theory, methods and
applications. Benjamins, Amsterdam, Philadelphia, 1999.</li>

        <li> Mayer, Felix/Seewald-Heeg, Uta (Hrsg.):
Terminologiemanagement. Von der Theorie zur Praxis. Berlin: BD&Uuml;, 2009.</li>

        <li> Mayer, Felix, Schmitz, Klaus-Dirk, Zeumer, Jutta (Hrsg.):
Terminologie und Wissensmanagement. Akten des Symposiums, Köln, 26.-27.
März 2004, Deutscher Terminologie-Tag e.V., 2004.</li>

        <li> Mayer, Felix, Schmitz, Klaus-Dirk, Zeumer, Jutta (Hrsg.):
eTerminology. Professionelle Terminologearbeit im Zeitalter des Internet. Akten
des Symposiums, Köln, 12.-13. April 2002, Deutscher Terminologie-Tag e.V.,
2002.</li>

        <li> Wright, Sue Ellen/Budin, Gerhard (1997/2001): Handbook of
Terminology Management. Vol. 1: Basic Aspects of Terminology Management. Vol. 2:
Application Oriented Terminology Management. Benjamins, Amsterdam, Philadelphia,
1997 und 2001. </li>

        <li> Terminologearbeit Best Practices. Deutscher
Terminologietag/deutsches Institut für Terminologie, 2010.</li>
    </ul>
</body>
</html>

```

Für das Deckblatt wurde eine Grafik mit dem Icon der Hochschule-Anhalt genommen, die in der Form einer JPEG-Datei vorhanden ist. Die Grafik soll mittels des aus HTML bekannten Inline-Elements `` markiert und mit den Attributen `src` und `alt` verwiesen und dargestellt werden. Bei der Positionierung aller Texte und Grafiken ist das Attribut `style` zuständig. Die Grafik wird außerdem durch manche E-Book-Reader als einen anzuzeigenden Umschlag dargestellt. Das heißt, dass diese Grafik mit Buchtitel in manchen digitalen Buchregalen angezeigt wird.

Quellcode 9. Deckblatt mit Stylesheet⁷³

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"

```

⁷³ Vgl. Data2type (o.J.): ePUB. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

```

"http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en">
  <head>
    <title> Modulhandbuch</title>
    <meta http-equiv="Content-Type" content="text/html; charset=gb2312"/>
    <style type="text/css" xml:space="preserve">
      table { max-width: 600px; width: 100%}
      body{font-family:Calibri}
      h1 {font-family:Calibri}
      li {line-height: 30px}
      body{margin-top:10%}
    </style>
  </head>
  <body>
    

    <h1>Modulhandbuch </h1>
    <hr style="color:#987cb9"/>
    <p style="color:#FF0000"> Hochschule Anhalt </p>
    <p style="color:#FF0000"> Fachbereich Informatik und Sprachen</p>
    <p style="display:block;position:absolute;bottom:10%; ">
Softwarelokalisierung</p>
  </body>
</html>

```

Um dieses EPUB-Dokument benutzerfreundlicher zu machen, wurde ein Button hinzugefügt, damit man ohne Blättern zur ersten Seite springen kann.

Quellcode 10. Button in der HTML-Datei⁷⁴

```

<button onclick="window.open('./deckblatt.xhtml', '_self', 'resizable=yes')">Zurück
zur ersten Seite</button>

```

Bei der Wahl von Datei- und Verzeichnisnamen gibt es nicht viele Beschränkungen. Im Allgemein müssen sie den Anforderungen von Betriebssystemen entsprechen.

⁷⁴ Vgl. Data2type (o.J.): ePUB. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

Dennoch ist es empfehlenswert, dass die Namen in einem Verzeichnis aufgrund der Kompatibilität sich nicht nur durch Klein- und Großschreibung unterscheiden. Daraufhin sollten die Namen komplett auf HTML-Entitäten und andere Sonderzeichen verzichten.

4.3.3 Darstellung und Kodierung von Schriften

Bücher im Format EPUB sind normalerweise mit UTF-8 oder UTF-16 codiert und mit CSS formatiert. Chinesisch wird oft ohne Regelung mit Garbled angezeigt. Manuelle Modifizierungen sind dafür notwendig. Laut IDPF enthält EPUB 3 CSS-Profil `@font-face`-Regeln und Deskriptoren. Dazu müssen Lesesysteme mit einen CSS-Viewport Opentype [Opentype] und WOFF [WOFF] Schriften unterstützen.

„OpenType ist ein modernes Datenformat für digitalisierte Satzschriften (Fonts), ursprünglich eingeführt von *Microsoft* (als TrueType Open) und später gemeinsam mit *Adobe* weiter entwickelt. Das OT-Format überwindet die Begrenzungen der weit verbreiteten Font-Formate PostScript (Type1) und TrueType, zum Beispiel den limitierten Zeichenvorrat“ [FontShop 2015]. Um ein Dokument in einer besonderen chinesischen Schriftart zu zeigen, muss eine Schriften-Datei im Archiv enthalten sein, die aus dem Internet heruntergeladen oder selbst erstellt werden kann.

4.4 Erstellung einer EPUB-Datei

Im vorherigen Kapitel wurden die Standards und die Dokumentstruktur von EPUB beschrieben. In diesem Kapitel wird erklärt, wie man ein EPUB-Dokument, genauer gesagt das Softwarelokalisierungs-Modulhandbuch, erstellen kann. Wie in den vorherigen Abschnitten bereits erläutert wurde, sind neben den (X)HTML-Dateien, die die Inhaltsdateien bilden, auch noch XML-Dateien, welche die Strukturen und Besonderheiten beschreiben, notwendig.

4.4.1 MIME-Type-Datei⁷⁵

Zuerst muss eine XML-Datei mit dem Namen *mimetype* erstellt werden. Diese Datei zeigt, dass es sich bei dem vorliegenden Dokument um eine EPUB-Publikation handelt. Die MIME-Type-Datei enthält nur eine Zeile:

```
application/epub +zip.
```

4.4.2 META-INF-Ordner mit *container.xml*⁷⁶

Danach wird das Verzeichnis META-INF gebaut. In diesem Verzeichnis wird die Datei *container.xml* angelegt und notwendige Inhalte eingefügt. An dieser Stelle wird der Rootfile-Pfad als "OPS/content.opf" vordefiniert. Der Quellcode 11 zeigt den Inhalt der Container-Datei.

Quellcode 11. Container-Datei⁷⁷

```
<?xml version="1.0" encoding="UTF-8"?>
<container version="1.0" xmlns="urn:oasis:names:tc:opendocument:xmlns:container">
  <rootfiles>
    <rootfile full-path="OPS/content.opf" media-type="application/oebps-
package+xml"/>
  </rootfiles>
</container>
```

4.4.3 OPS-Verzeichnis⁷⁸

Im nächsten Schritt muss ein weiteres Verzeichnis erstellt werden. Das OPS-Verzeichnis wird aus einer Reihe von XHTML- und XML-Dateien und noch weiteren Unterverzeichnissen bestehen. Sämtliche in vorherigen Kapiteln erwähnte und

⁷⁵ Siehe Abschnitt 3.5.2

⁷⁶ Siehe Abschnitt 3.5.3

⁷⁷ Vgl. Data2type (o.J.): ePUB. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

⁷⁸ Siehe Abschnitt 3.5.4

vorbereitete Dateien werden in dieses Verzeichnis bzw. in das zugehörige Unterverzeichnis kopiert.

4.4.3.1 OPS / Datei content.opf

Für die einzigartige ID *unique-identifizier* wird beispielsweise der Wert "12345-bookid-2015" angegeben. Zusätzlich ist noch die Sprache innerhalb des <dc:language>-Elementes für die Lokalisierung von Bedeutung. Außer dem Deckblatt werden alle weiteren Inhaltsdateien mit Defaultwert („yes“) die lineare Lesefolge beibehalten.

Quellcode 12. Struktur der OPF-Datei⁷⁹

```
<?xml version='1.0' encoding='utf-8'?>
<!DOCTYPE package PUBLIC "-//ISBN 978-7-308-05831-5//DTD OEB 1.2 Package//EN"
"http://openebook.org/dtds/oeb-1.2/oebpkg12.dtd">
<package unique-identifizier="12345-bookid-2015"
xmlns="http://www.idpf.org/2007/opf" xmlns:dc="http://purl.org/dc/elements/1.1/"
version="2.0">
<metadata xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:opf="http://www.idpf.org/2007/opf">
  <dc:title>Modulhandbuch-Softwarelokalisierung</dc:title>
  <dc:identifizier opf:scheme="ISBN">12345-bookid-2015</dc:identifizier>
  <dc:language>de</dc:language>
  <dc:creator>Hochschule Anhalt</dc:creator>
  <dc:subject>Modulhandbuch</dc:subject>
  <dc:contributor> Wang Yaochen</dc:contributor>
  <dc:language>de</dc:language>
  <dc:publisher/>
  <dc:description/>
  <dc:coverage/>
  <dc:source/>
  <dc:date>2015</dc:date>
  <dc:rights/>
  <dc:subject>Modulhandbuch</dc:subject>
  <dc:contributor/>
  <dc:type>[type]</dc:type>
```

⁷⁹ Vgl. Data2type (o.J.): ePUB. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

```
<dc:format/>
<dc:relation/>
<meta name="cover" content="cover-image" />
</metadata>
</metadata>
<manifest>
  <item id="cover" href="Deckblatt.xhtml" media-
type="application/xhtml+xml"/>
  <item id="Inhaltsverzeichnis" href="Inhaltsverzeichnis.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul01" href="modul01.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul02" href="modul02.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul03" href="modul03.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul04" href="modul04.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul05" href="modul05.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul06" href="modul06.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul07" href="modul07.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul08" href="modul08.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul09" href="modul09.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul10" href="modul10.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul11" href="modul11.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul12" href="modul12.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul13" href="modul13.xhtml" media-
type="application/xhtml+xml"/>
  <item id="modul14" href="modul14.xhtml" media-
type="application/xhtml+xml"/>
```

```
        <item id="modul15" href="modul15.xhtml" media-
type="application/xhtml+xml"/>
        <item id="modul16" href="modul16.xhtml" media-
type="application/xhtml+xml"/>
        <item id="cover-image" href="images/coverimage.jpg" media-
type="image/jpg"/>
        <item id="ncx" href="toc.ncx" media-type="application/x-dtbncx+xml"/>
        <item id="css" href="ebook.css" media-type="text/css"/>
    </manifest>
<spine toc="ncx">
    <itemref idref="cover" linear="no"/>
    <itemref idref="Inhaltsverzeichnis" linear="no"/>
    <itemref idref="modul01"/>
    <itemref idref="modul02"/>
    <itemref idref="modul03"/>
    <itemref idref="modul04"/>
    <itemref idref="modul05"/>
    <itemref idref="modul06"/>
    <itemref idref="modul07"/>
    <itemref idref="modul08"/>
    <itemref idref="modul09"/>
    <itemref idref="modul10"/>
    <itemref idref="modul11"/>
    <itemref idref="modul12"/>
    <itemref idref="modul13"/>
    <itemref idref="modul14"/>
    <itemref idref="modul15"/>
    <itemref idref="modul16"/>
</spine>
<guide>
    <reference href="Modulhandbuch-Deckblatt.html" type="cover" title="Cover"/>
</guide>
</package>
```

4.4.3.2 OPS / Datei toc.ncx bzw. Datei nav.xhtml

Zur linearen Lesereihenfolge des Lesens dient die NCX-Datei, die hier als ein Inhaltsverzeichnis funktioniert. Ein Teil der in der OPF-Datei beschriebenen Informationen wird hier wiederholt. Die Lehrveranstaltungen für den Masterstudiengang Softwarelokalisierung können sowohl nach einer alphabetischen Reihenfolge als auch chronologisch geordnet werden. Jedoch ist eine gemischte Reihenfolge nach Zeit und Logik für die Studierenden sinnvoller. Die Modulnamen werden also zunächst in die dazugehörigen Semester eingeteilt. Dann werden die Pflichtmodule vor die Wahlpflichtmodule gestellt. Schließlich werden die unterteilten Modulnamen nach der alphabetischen Regel geordnet. Die Datei *toc.ncx* dient als das Inhaltsverzeichnis von EPUB 2 und ist nicht komplex. Im Quellcode 13 werden die Dateiinhalte formuliert (Siehe Quellcode 13).

Quellcode 13. Struktur der Inhaltsverzeichnis-Datei von NCX⁸⁰

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ncx PUBLIC
    "-//NISO//DTD ncx 2005-1//EN"
    "http://www.daisy.org/z3986/2005/ncx-2005-1.dtd">
<ncx xmlns="http://www.daisy.org/z3986/2005/ncx/" version="2005-1" xml:lang="de"
dir="ltr">
    <head>
        <meta name="dtb:uid" content="12345-bookid-2015"/>
        <meta name="dtb:depth" content="2"/>
        <meta name="dtb:totalPageCount" content="0"/>
        <meta name="dtb:maxPageNumber" content="0"/>
    </head>
    <docTitle>
        <text> Modulhandbuch-Softwarelokalisierung </text>
    </docTitle>
    <docAuthor>
        <text>Hochschule Anhalt Softwarelokalisierung</text>
    </docAuthor>
    <navMap>
```

⁸⁰ Vgl. Data2type (o.J.): ePUB. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

```

<navPoint id="Inhaltsverzeichnis" playOrder="1">
  <navLabel>
    <text>Inhaltsverzeichnis</text>
  </navLabel>
  <content src="Inhaltsverzeichnis.xhtml"/>
</navPoint>
<navPoint id="modul02" playOrder="2">
  <navLabel>
    <text>Effizientes Terminologiemangement</text>
  </navLabel>
  <content src="../ops/dokument.htm#modul02"/>
  <navPoint id="modul03" playOrder="3">
    <navLabel>
      <text>IT- und translationswissenschaftliche
Grundlagen der Softwarelokalisierung</text>
    </navLabel>
    <content src="../ops/dokument.htm#modul03"/>
  </navPoint>
  <navPoint id="modul04" playOrder="4">
    <navLabel>
      <text>Lokalisierung grafischer
Benutzungsoberflächen</text>
    </navLabel>
    <content src="../ops/dokument.htm#modul04"/>
  </navPoint>
  <navPoint id="modu05" playOrder="5">
    <navLabel>
      <text>Übersetzungs- Lokalisierungstechnologie
Austauschformate</text>
    </navLabel>
    <content src="../ops/dokument.htm#modul05"/>
  </navPoint>
  <navPoint id="modul06" playOrder="6">
    <navLabel>
      <text>Angewandte Sprachwissenschaft</text>
    </navLabel>
    <content src="../ops/dokument.htm#modul06"/>
  </navPoint>
  <navPoint id="modul07" playOrder="7">

```

```
<navLabel>
  <text>Internationales Marketing</text>
</navLabel>
<content src="../ops/dokument.htm#modul07"/>
</navPoint>
<navPoint id="modul08" playOrder="8">
  <navLabel>
    <text>Softwareentwicklung und
Internationalisierung</text>
  </navLabel>
  <content src="../ops/dokument.htm#modul08"/>
</navPoint>
<navPoint id="modul09" playOrder="9">
  <navLabel>
    <text>Übersetzen Lokalisierung produktbegleitender
Texte</text>
  </navLabel>
  <content src="../ops/dokument.htm#modul09"/>
</navPoint>
<navPoint id="modul10" playOrder="10">
  <navLabel>
    <text>XML</text>
  </navLabel>
  <content src="../ops/dokument.htm#modul10"/>
</navPoint>
<navPoint id="modul11" playOrder="11">
  <navLabel>
    <text>Lokalisierungsprojekt (Master)</text>
  </navLabel>
  <content src="../ops/dokument.htm#modul12"/>
</navPoint>
<navPoint id="modul12" playOrder="12">
  <navLabel>
    <text>Projektmanagement in
Lokalisierungsprojekten</text>
  </navLabel>
  <content src="../ops/dokument.htm#modul13"/>
</navPoint>
<navPoint id="modul13" playOrder="13">
```

```

        <navLabel>
            <text>Qualitätssicherung in
Lokalisierungsprojekten</text>
        </navLabel>
        <content src="../../ops/dokument.htm#modul13"/>
    </navPoint>
</navPoint>
<navPoint id="modul14" playOrder="14">
    <navLabel>
        <text>Übersetzungsgerechtes Schreiben</text>
    </navLabel>
    <content src="../../ops/dokument.htm#modul014"/>
</navPoint>
<navPoint id="modul15" playOrder="15">
    <navLabel>
        <text>Lokalisierung von Grafik und Multimedia</text>
    </navLabel>
    <content src="../../ops/dokument.htm#modul15"/>
</navPoint>
<navPoint id="modul16" playOrder="16">
    <navLabel>
        <text>Technisches Schreiben für Online-Medien</text>
    </navLabel>
    <content src="../../ops/dokument.htm#modul16"/>
</navPoint>
<navPoint id="modul17" playOrder="17">
    <navLabel>
        <text>Übersetzen technischer Fachtexte außerhalb des IT-
Bereichs</text>
    </navLabel>
    <content src="../../ops/dokument.htm#modul17"/>
</navPoint>
</navMap>
</ncx>

```

Für die EPUB Version 3 wird das NCX-Format des Inhaltsverzeichnisses durch ein einfaches XHTML-Format ersetzt. Alle Titelnamen innerhalb des Inhaltsverzeichnisses sind unter dem Element innerhalb des Unterelements

untergebracht. Jeder Punkt des Inhaltsverzeichnisses wird durch das Attribut `a` eingeschränkt (Siehe Quellcode 14).

Quellcode 14. XHTML-Inhaltsverzeichnis-Datei-Struktur ⁸¹

```
<?xml version="1.0" encoding="UTF-8"?>
<html xmlns="http://www.w3.org/1999/xhtml"
xmlns:ops="http://www.idpf.org/2007/ops" xml:lang="de">
  <head>
    <style type="text/css" xml:space="preserve">
      table { max-width: 600px; width: 100%}
      body{font-family:Calibri}
      h1 {font-family:Calibri}
      li {line-height: 30px}
    </style>
    <title>Modulhandbuch</title>
  </head>
  <body>
    <h1>Modulhandbuch<br/>
    </h1>
    <nav ops:type="toc">
    <h2>Inhaltsverzeichnis</h2>
    <ol>
      <li>
        <a
href="Inhaltsverzeichnis.xhtml">Inhaltsverzeichnis</a>
      </li>
      <li>
        <a href="modul01.xhtml">Effizientes
Terminologiemangement</a>
      </li>
      <li>
        <a href="modul02.xhtml">IT- und
translationswissenschaftliche Grundlagen der Softwarelokalisierung</a>
      </li>
      <li>
        <a href="modul03.xhtml">Lokalisierung grafischer
Benutzungsoberfl&auml;chen</a>

```

⁸¹ Vgl. Data2type (o.J.): ePUB. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

```
</li>
<li>
    <a href="modul04.xhtml">&Uuml;bersetzungs-
Lokalisierungstechnologie Austauschformate</a>
</li>
<li>
    <a href="modul05.xhtml">Angewandte
Sprachwissenschaft</a>
</li>
<li>
    <a href="modul06.xhtml">Internationales Marketing</a>
</li>
<li>
    <a href="modul07.xhtml">Softwareentwicklung und
Internationalisierung</a>
</li>
<li>
    <a href="modul08.xhtml">&Uuml;bersetzen Lokalisierung
produktbegleitender Texte</a>
</li>
<li>
    <a href="modul09.xhtml">XML</a>
</li>
<li>
    <a href="modul10.xhtml">Lokalisierungsprojekt
(Master)</a>
</li>
<li>
    <a href="modul11.xhtml">Projektmanagement in
Lokalisierungsprojekten</a>
</li>
<li>
    <a href="modul12.xhtml">Qualit&auml;tssicherung in
Lokalisierungsprojekten</a>
</li>
<li>
    <a href="modul13.xhtml">&Uuml;bersetzungsgerechtes
Schreiben</a>
</li>
```

```
        <li>
            <a href="modul14.xhtml">Lokalisierung von Grafik und
Multimedia</a>
        </li>
        <li>
            <a href="modul15.xhtml">Technisches Schreiben f&uuml;r
Online-Medien</a>
        </li>
        <li>
            <a href="modul16.xhtml">&Uuml;bersetzen technischer
Fachtexte au&szlig;erhalb des IT-Bereichs</a>
        </li>
    </ol>
</nav>
</body>
</html>
```

Die Inhaltsverzeichnis-Datei im XHTML-Format und die Inhaltsverzeichnis-Datei von NCX stehen im EPUB-Dokument alternativ zur Auswahl.

4.4.3.3 OPS / Datei images / ...

Alle in den Dateien auftauchenden Grafiken sollen im Unterverzeichnis *images* liegen. Zu beachten ist, dass jeder Name der Grafiken mit dem in den anderen Dateien geschriebenen Namen bzw. IDs übereinstimmen muss. Unterschiedliche Programme stellen verschiedene Aussehen eines E-Books dar. Im digitalen Buchregal kann ein E-Book mit einem Deckblatt oder der ersten Seite im E-Book-Reader angezeigt werden. Die Datei *cover.jpg* bietet für manche Programme ein Deckblatt an. Die Datei *coverimage.jpg* wird im Deckblatt einbezogen.

4.4.3.4 OPS / Datei CSS / Modulandbuch.css

Eventuell sollte ein Unterverzeichnis für Stylesheets erstellt werden. Eine CSS-Datei kann wie folgt formuliert sein (Siehe Quellcode 15):

Quellcode 15. Stylesheet-Datei-Struktur⁸²

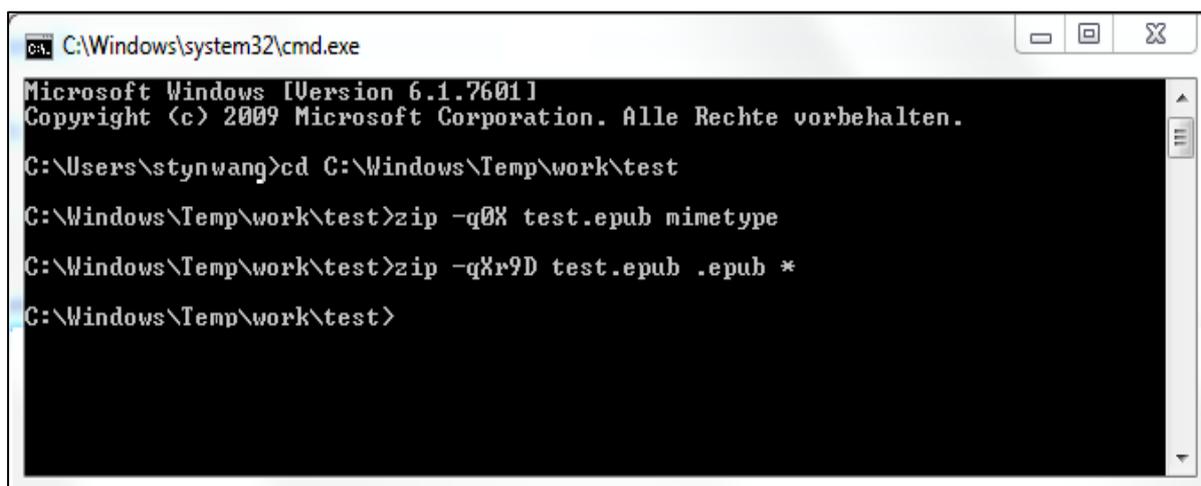
```
body {
    font-family: constantia;
    padding: 0%;
    margin-top: 0%;
    margin-bottom: 0%;
    margin-left: 1%;
    margin-right: 1%;
    line-height:130%;
    text-align: justify;
    font-family:"cnepub";
}
div {
    margin:0px;
    padding:0px;
    line-height:130%;
    text-align: justify;
    font-family:"cnepub";
}
p {
    text-align: justify;
    text-indent: 0px;
    line-height:130%;
    margin-bottom: -0.8em;
}
.cover {
    width:100%;
    padding:0px;
}
h1 {
    font-family:Calibri;
    line-height:130%;
    text-align: left;
    font-weight:bold;
    font-size:xx-large;
```

⁸² Vgl. Data2type (o.J.): Die Metadaten der Publikation. Online im Internet:
<http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

4.5 Erstellung eines ZIP-Archivs

Am Anfang oder am Ende der Arbeit wird ein leeres ZIP-Archiv erstellt. Weil die Datei „mimetype“ weder komprimiert noch verschlüsselt werden darf, und als die erste Datei im Archiv gespeichert werden soll, funktioniert ein einfaches Zippen aller Dateien an dieser Stelle nicht. Es ist möglich, dass eine Datei unter dem Namen „Modulhandbuch_Master-Softwarelokalisierung“ mit der Erweiterung .zip gespeichert wird. In dieses Archiv muss die MIME-Type-Datei unkomprimiert zu Beginn gespeichert werden. Anschließend können die anderen Verzeichnisse auch komprimiert hinzugefügt werden. Somit sind dann sämtliche Teildateien in einem Zip-Archiv zusammengefasst. Abschließend muss nun nur noch die Dateinamenserweiterung von .zip zu .epub geändert werden. Zu beachten ist, dass EPUB die Verschlüsselung beim Zippen oder bei der MIME-Type-Datei nicht unterstützt.⁸³

Darüberhinaus ist es auch möglich, eine EPUB-Datei durch Eingabe im Terminal zu erstellen. Es muss darauf geachtet werden, dass die MIME-Type-Datei als erste Datei im Zip-Archiv erscheint. In der Kommandozeile im Terminal müssen die folgenden Befehle eingegeben werden (Siehe Abbildung 10).

The image shows a screenshot of a Windows Command Prompt window. The title bar reads "C:\Windows\system32\cmd.exe". The window content shows the following text:

```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. Alle Rechte vorbehalten.

C:\Users\stynwang>cd C:\Windows\Temp\work\test
C:\Windows\Temp\work\test>zip -q0% test.epub mimetype
C:\Windows\Temp\work\test>zip -q%r9D test.epub .epub *
C:\Windows\Temp\work\test>
```

Abbildung 10. Erzeugen eines Zip-Archivs im Terminal

⁸³ Vgl. Data2type (o.J.): Die Metadaten der Publikation. Online im Internet: <http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/> [09.09.2015]

4.6 Überprüfung

Die erzeugten Dateien können mit unterschiedlichen Programmen auf verschiedenen Betriebssystemen überprüft werden, um zu sehen, ob es Probleme bei der Programmierung oder bei den Inhalten gibt. Die einzelnen Dateien müssen den standardisierten Regeln von XML und (X)HTML entsprechen. Außerdem müssen alle Regeln des EPUB-Standards eingehalten werden.

Um das erzeugte EPUB-Dokument zu testen, sind spezielle Programme nötig. Dafür wurden in diesem Projekt auf dem PC die Software Adobe Digital Editions und auf dem iPad-Gerät die Software ibooks verwendet. Ein Dokument kann durch verschiedene Programme unterschiedliche Ergebnisse liefern. Als Beispiele zeigen die folgenden Abbildungen die Darstellungen des Modulhandbuchs der Fachrichtung Softwarelokalisierung von Adobe Digital Editions und ibooks.

Zunächst wird geprüft, ob das EPUB-Dokument im E-Book-Reader geöffnet werden kann. Falls es nicht funktioniert oder nicht vollständig angezeigt werden kann, besteht sicherlich ein Fehler an dieser Stelle. Oftmals treten gleiche oder zumindest ähnliche Probleme an ähnlichen Stellen auf. Das fehlerhafte Dokument wird dann wieder korrigiert, bis alle Inhalte, sowohl Texte als auch Grafiken und andere Multimedia-Daten, angezeigt werden. Die Inhalte wurden direkt aus den Word-Dateien extrahiert, sie sollen insofern keine Diskrepanzen zwischen originalen und transportierten Inhalten aufweisen. Gleichzeitig werden die Reihenfolge aller Kapitel und E-Book-Funktionen wie Durchblättern getestet. Schließlich kommt das Korrekturlesen von allen Inhalten vor. Allerdings können nicht alle Probleme automatisch identifiziert werden. Es gibt bisher kein Hilfsmittel, das den Tester auf die Fehlerstelle hinweist. Es muss daher stets eine manuelle Überprüfung und Korrektur durch den Menschen erfolgen.

Bei der Verwendung der Software Adobe Digital Editions wird das Inhaltsverzeichnis in dem Navigationsfenster links angezeigt und lässt sich ein- bzw. ausklappen. Auf der ersten Seite wird das Deckblatt nicht mit sämtlichen Elementen angezeigt (Siehe Abbildung11). Bei dem Test wird festgestellt, dass mit Adobe Digital Editions eine

Zeileninformation und die Grafik, die das Logo der Hochschule darstellt, nicht dargestellt werden.

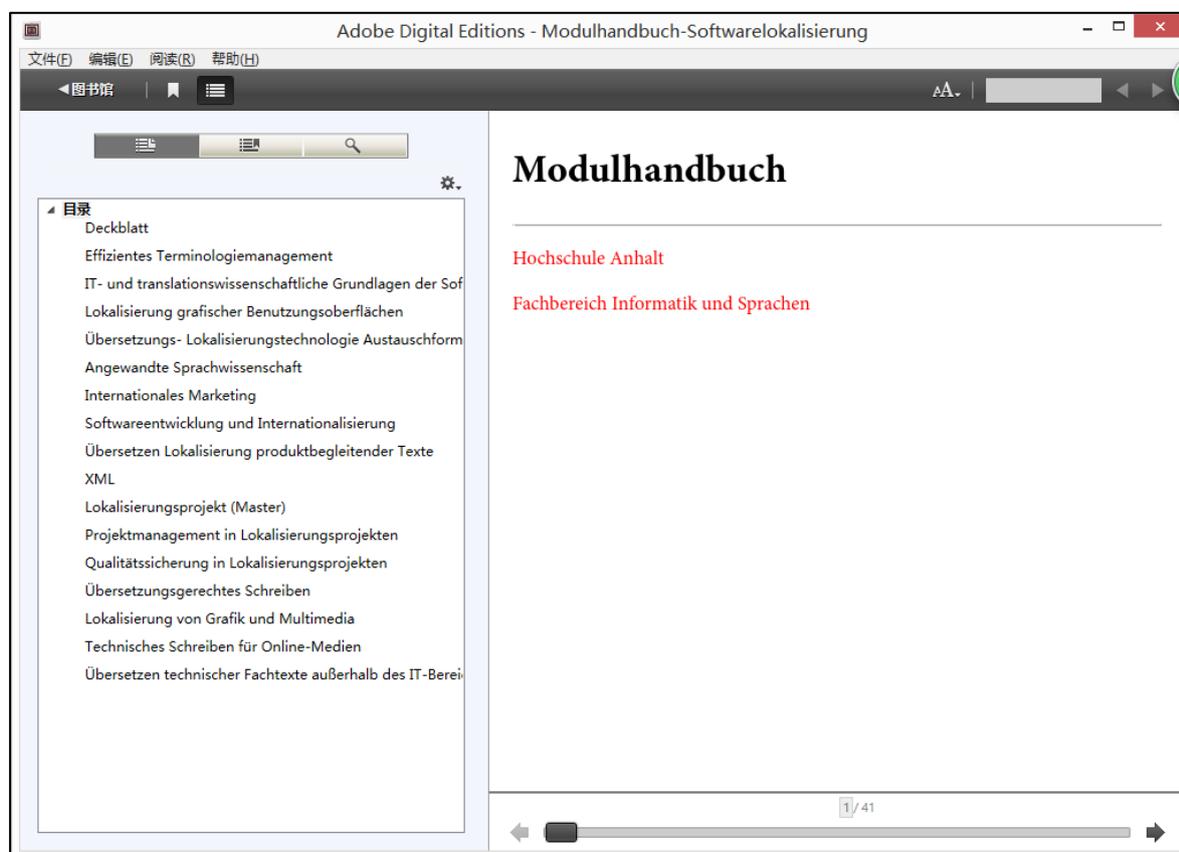


Abbildung 11. Test mit Adobe Digital Editions

Im Gegensatz zu Adobe Digital Editions liefert ibooks eine bessere Darstellung aller Details des EPUB-Dokuments. Der Dokumenttitel (Element <dc:titel>) und der Hersteller (Element <dc:creator>) werden abwechselnd auf der geraden und ungeraden Seite in der Kopfzeile angegeben. Die Grafik auf dem Deckblatt wird sogar perfekt präsentiert. (Siehe Abbildung 12)

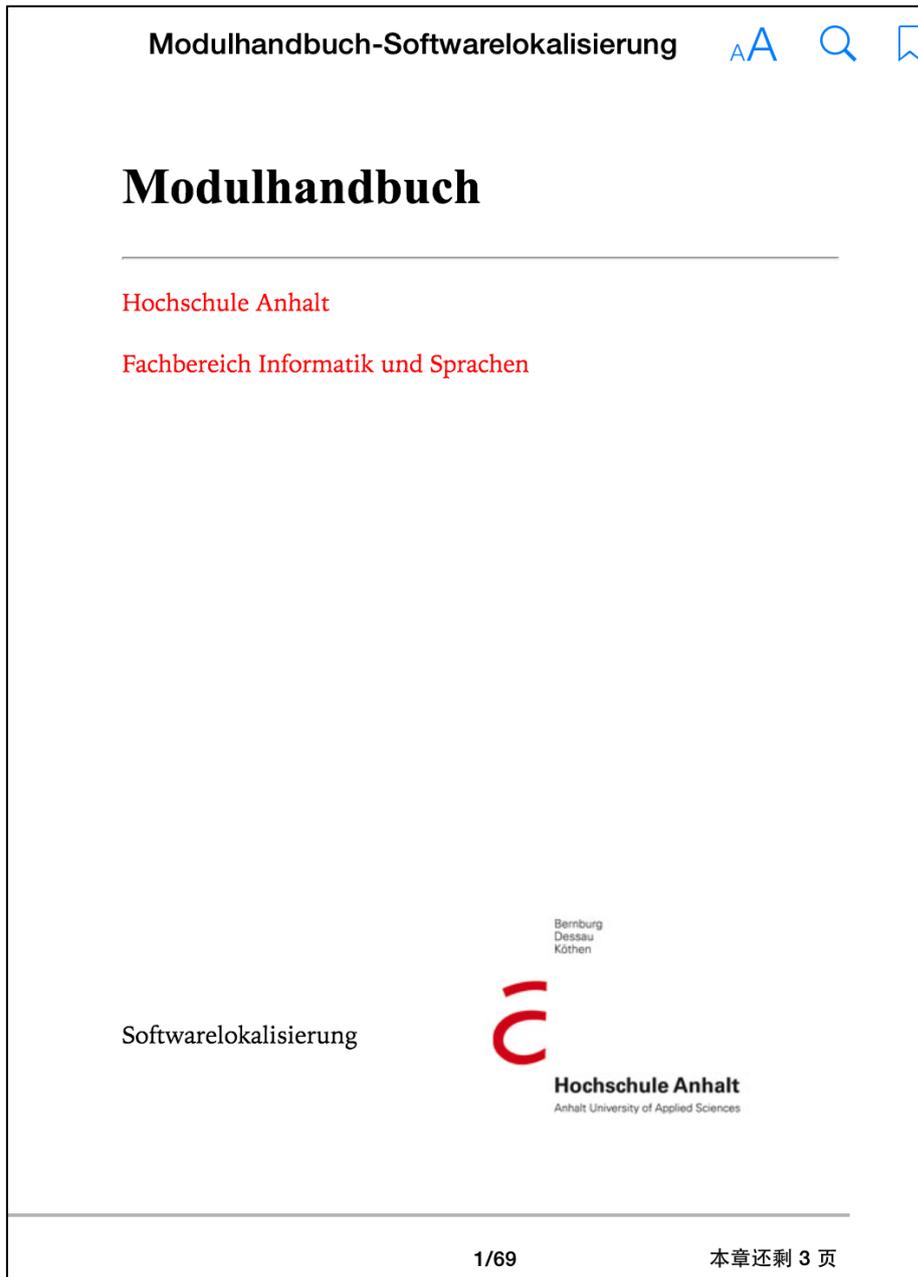


Abbildung 12. Deckblatt des EPUB-Dokuments in ibooks

Die als Deckblatt eingebettete Grafik funktioniert bei Adobe Digital Editions gar nicht, während ibooks diese gleich an zwei Stellen anzeigt. In der „Buchbibliothek“ steht dieses Bild neben dem Buchtitel und Autor als Vorschau des Buches. Wird das Inhaltsverzeichnis in ibooks geöffnet, wird die Grafik auf der linken Seite wiedergegeben, wenn das Dokument quer auf dem Bildschirm angezeigt wird (Siehe Abbildung 13).

Modulhandbuch-Softwarelokalisierung		
Inhalte	Lesezeichen	Notizen
Deckblatt		1
Effizientes Terminologiemanagement		2
IT- und translationswissenschaftliche Grundlagen der Softwarelokalisierung		6
Lokalisierung grafischer Benutzungsoberflächen		10
Übersetzungs- Lokalisierungstechnologie Austauschformate		15
Angewandte Sprachwissenschaft		20
Internationales Marketing		24
Softwareentwicklung und Internationalisierung		28
Übersetzen Lokalisierung produktbegleitender Texte		33
XML		37
Lokalisierungsprojekt (Master)		40
Projektmanagement in Lokalisierungsprojekten		44

Abbildung 13. Quer-Anzeige in ibooks

Bis alles in Ordnung ist, kann das Dokument auf anderen E-Book-Readern bzw. mit anderen Programmen verwendet werden, sodass die EPUB-Anwendung als funktionstüchtig gelten kann. Durch diese Schritte wird die Qualität des erzeugten EPUB-Dokuments gewährleistet. Auf weitere Korrekturen und Regelungen der Inhalte wird in Kapitel 5 eingegangen.

The screenshot shows the Adobe Digital Editions interface. The window title is 'Adobe Digital Editions - Modulhandbuch-Softwarelokalisierung'. The top menu bar includes '文件(F)', '编辑(E)', '阅读(R)', and '帮助(H)'. The left sidebar contains a table of contents with the following items:

- 目录
- Deckblatt
- Effizientes Terminologiemanagement
- IT- und translationswissenschaftliche Grundlagen der Softwarelokalisierung
- Lokalisierung grafischer Benutzeroberflächen
- Übersetzungs- Lokalisierungstechnologie Austauschform
- Angewandte Sprachwissenschaft
- Internationales Marketing
- Softwareentwicklung und Internationalisierung
- Übersetzen Lokalisierung produktbegleitender Texte
- XML
- Lokalisierungsprojekt (Master)
- Projektmanagement in Lokalisierungsprojekten
- Qualitätssicherung in Lokalisierungsprojekten
- Übersetzungsgerechtes Schreiben
- Lokalisierung von Grafik und Multimedia
- Technisches Schreiben für Online-Medien
- Übersetzen technischer Fachtexte außerhalb des IT-Bereichs

The main content area displays the following information for the module 'Effizientes Terminologiemanagement':

Effizientes Terminologiemanagement

Modulbezeichnung: Effizientes Terminologiemanagement
Zuordnung: Pflichtmodul für Softwarelokalisierung
Semester: 1. Fachsemester
verantwortlich: Prof. Dr. Uta Seewald-Heeg
Dozent / Dozentin: Prof. Dr. Uta Seewald-Heeg, Valentina Uswak M.A.
Sprache: deutsch
Credits: 5

Modulumfang (in SWS)	Arbeitsaufwand (in Stunden)
Vorlesung: 1	Präsenzstudium: 45
Übung: 1	Eigenstudium: 105
Praktikum: 1	

Modulziele / Angestrebte Lernergebnisse:

Die Studierenden erwerben Wissen über den Aufbau und die Struktur von Terminologieverwaltungssystemen (TVS). Sie verstehen die strukturellen Konzepte von TVS und haben die Fähigkeit, Ressourcen für die Extraktion von Terminologie unterschiedlicher Sachgebiete zu ermitteln und qualitativ zu

The page number '2 / 40' is visible at the bottom of the page.

Abbildung 14. Buchinhalt in Adobe Digital Editions

5. Lokalisierung des erzeugten EPUB-Dokuments

In diesem Kapitel geht es um die umfangreichen Lokalisierungsprozesse des erzeugten EPUB-Dokuments mittels des Lokalisierungsprogramms SDL Trados Studios. Die Schwerpunkte werden präzisiert. Dabei werden ausführliche Arbeitsschritte nicht aufgezählt.

5.1 Lokalisierungsprojekt

Mit der stabilen nachhaltigen Entwicklung der Beziehung zwischen Deutschland und China in vielen Aspekten erstellen die beiden Länder eine offene Plattform für wirtschaftliche und akademische Zusammenarbeit und Kommunikation. Durch die Politikreform und -regelung lernen immer mehr Chinesen Deutschland und seine hervorragende Infrastruktur, hohe Lehrqualität und niedrige Ausbildungskosten kennen. Infolgedessen haben sie die Absicht, nach Deutschland zu fliegen und dort theoretische und praktische Kenntnisse zu erwerben.

Die Statistiken zeigen, dass die Anzahl der chinesischen Studierenden in Deutschland Jahr für Jahr zunimmt. 2013 betragen die an deutschen Hochschulen immatrikulierten chinesischen Studenten insgesamt 25.564. Im Vergleich zum Vorjahr ist die Anzahl um 7% gestiegen. Die chinesischen Studierenden bilden die größte Gruppe der ausländischen Studierenden im ganzen Land.⁸⁴

Allerdings ist die Fremdsprache immer noch ein großes Hindernis für den Zugriff und das Verstehen der Informationen. Für die Bewerber sind die Vorbereitung und das Anbieten aller Informationen der Hochschule von Bedeutung. Um die Interessierenden mit der deutschen Sprache und Kultur vertraut zu machen, spielt die Lokalisierung der Materialien als Propagandamittel eine große Rolle. Das ist der wesentliche Grund für die steigende Anforderung der interkulturellen Kommunikation und der zunehmenden Nachfrage zur Lokalisierung von Dokumentationen.

⁸⁴ Vgl. 2014 年度中国留学人员统计 (Jährliche chinesischen Studenten-Statistik 2014) http://www.demoe.edu.cn/article_read.php?id=12014-20141216-2145

Wie im Abschnitt 4.1 schon erläutert wurde, ist die Lokalisierung ein Teil von GILT. Der GILT-Zyklus stellt die Unternehmen vor die Herausforderungen, nach denen diese ihre Produkte möglichst umfassend und breit an die Zielländer oder die Zielmärkte anpassen, um reißenden Absatz zu finden. Die Internationalisierung eines Produktes erleichtert die Tätigkeiten der Lokalisierung an einen lokalen Markt und vermindert demzufolge die Zeit und Kosten der Weiterentwicklung.

Die Lokalisierung ist „die Anpassung eines Software-Produkts an einen lokalen Markt mit seinen sprachlichen und kulturellen Besonderheiten mit Hilfe spezieller technologischer Methoden und Werkzeuge“ [Seewald-Heeg 2009]. Die Lokalisierung entspricht nicht gleich der Übersetzung. Sie umfasst nicht nur die Übersetzung von reinen Texten, sondern auch die Anpassung der formellen, kulturellen und sonstigen Anforderungen eines bestimmten Landes bzw. Zielmarktes. Bei der Translation bzw. der Übersetzung wird von der semantischen und äquivalenten Übertragung von schriftlichen Inhalten sowie kulturbezogenen Bestandteilen ausgegangen. Die Übersetzungsprozesse sind oft mit der Lokalisierung von Datum, Maßeinheit, Währungseinheit usw. verbunden.

Ziel des Projektes ist es, das erzeugte EPUB-Dokument bezüglich der Modulinformationen von Master Softwarelokalisierung vom Deutschen ins Chinesische zu lokalisieren.

Ein EPUB-Dokument ist eine mobile Anwendung, die alle Webdateien und Stilvorlagendateien beinhaltet. Die Lokalisierung von EPUB-Dokumenten basiert grundsätzlich auf der Webseite-Lokalisierung, weil sämtliche zu übersetzenden Inhalte Texte innerhalb des Start- und End-Tags in einer XHTML- bzw. XML-Datei sind. Das Erkennen der von der Codierung in der Lokalisierungssoftware getrennten Inhalte spielt bei der Lokalisierung daher eine bedeutende Rolle. Das erleichtert nicht nur die Übersetzungsarbeit, sondern vermeidet auch die Gefahr, die Quellkodierung des Produktes zu ändern.

5.2 Lokalisierungstechnologie

In diesem Projekt werden verschiedene Lokalisierungstechnologien zur Hilfe genommen. CAT ist das englische Akronym für **C**omputer **A**ided **T**ranslation oder auch **C**omputer **A**ssisted **T**ranslation, bzw. computergestützte Übersetzung im deutschen Sprachgebrauch genannt. Unter CAT versteht man eine menschliche Übersetzung mit Hilfe von Computerprogrammen mit dazugehörigen Funktionen. CAT verknüpft die menschliche Übersetzungsarbeit mit Übersetzungscomputerprogrammen. Der Begriff CAT ist von der MÜ (**M**aschinelle **Ü**bersetzung) zu differenzieren. MÜ liefert die Übersetzungsergebnisse im Gegensatz von CAT ohne menschliche Interferenz. CAT zeichnet sich meist durch hervorragende Qualität der Übersetzung und hohe Leistungsfähigkeit aus. CAT bietet ferner eine Wiederverwendung aller Datenbanken an. In erster Linie betrifft CAT ein Hilfsmittel TMS (**T**ranslation **M**emory **S**ystem).

Bei TM (**T**ranslation **M**emory) geht es um die Wiederverwendung von vorhergegangenen Übersetzungen. Dies ist sowohl in derselben Datei als auch im vollkommen anderen Projekt möglich. TM-Systeme fassen die Technologien von der Textverarbeitung und der Terminologieverwaltung zusammen. TMs sind Übersetzungsspeicher, sozusagen Datenbanken mit zweisprachigen bzw. mehrsprachigen Segmentpaaren, die im Laufe des Übersetzens verglichen und sogar erweitert werden können. Die Segmentpaare können satzweise gespeichert werden oder unvollständige Sätze sein. Jedes Segmentpaar unabhängig von der Länge wird als einzelner Eintrag, sozusagen eine Übersetzungseinheit, in der TM-Datenbank betrachtet. Man kann mit Hilfe eines Alignement-Programms ein Ausgangstext mit seinem vorliegenden ZIELTEXT verbinden und ein TM erzeugen. Während des Übersetzens werden die im TM gespeicherten Ausgangstexte mit dem zu übersetzenden Text verglichen. Erreicht der Übereinstimmungsgrad der beiden Texte die Prozentzahl aus der Einstellung, wird das gespeicherte Satzpaar abgerufen und im TM-Fenster als Übersetzungsvorschlag angezeigt. Der Übersetzer kann den Vorschlag direkt übernehmen oder als Referenz betrachten. Die mehr als 100% Übereinstimmungen (wie Perfekt Match und Context Match) können automatisch als Übersetzungen im ZIELTEXTFELD übernommen werden. Zum Beispiel

braucht man dadurch sich wiederholende Texte nicht mehr zu bearbeiten und spart somit nicht nur etwas Zeit, sondern hält das Ganze auch konsistent.

Eine Termbank hilft nicht nur bei der Vereinfachung der Übersetzungsarbeit, sondern auch beim Halten der Konsistenz aller Wörter.

5.3 Lokalisierungswerkzeuge

Im Lokalisierungsprozess sind Lokalisierungswerkzeuge die wichtigsten Hilfsmittel für Lokalisierer. Bei der Lokalisierung des Modulhandbuchs der Fachrichtung Softwarelokalisierung wurden die Lokalisierungsprodukte der Firma SDL verwendet, da sie auf dem Markt sehr verbreitet sind und zahlreiche Formate unterstützen.

SDL hat eine Reihe von Softwareprodukten entwickelt. Darunter sind SDL MultiTerm, SDL Trados Studio und SDL Passolo als die wesentlichen Produkte. SDL MultiTerm ist eine Terminologiedatenbank, die 1990 als das erste Produkt von SDL entwickelt wurde. MultiTerm dient als ein Terminologieverwaltungswerkzeug, um bilinguale bzw. multilinguale Terminologie zu verwalten. Für Terminologiedatenbanken können TBX oder andere proprietäre Formate wie SDLTB als Austauschformate je nach Präferenz verwendet werden. SDL Trados Studio ist ein Translation-Memory-System des britischen Unternehmens, das zu den hilfreichen CAT-Tools zählt. Im Jahr 2009 wurde Trados Studio in der SDL-Produktfamilie eingeschlossen. SDL Trados Studio ermöglicht mit Hilfe des TM (Translation Memory) sowohl die Übertragung der zu übersetzenden Inhalte als auch die Beibehaltung des Layouts der Software durch das Erkennen des Codes. SDL Passolo unterstützt bei der Lokalisierung von Benutzeroberflächen. Für Translation Memory sind TMX, XLIFF und proprietäre Formate wie SDLXLIFF als Austauschformate verwendbar.

Die Produkte von SDL Language Technologies können vielfältig im Lokalisierungsprozess eingesetzt werden. SDL-Produkte unterstützen zahlreiche Dateiformate, wie HTML, XML, MS Office-Dateien, InDesign, TBX usw. sowie eine große Zahl natürlicher Sprachen.

SDL Trados Studio ist ein integriertes Programm für das Übersetzen und die Lokalisierung. Trados Studio verfügt neben TM und Termbank noch über die Funktion der MÜ für Texte, bei denen keine Treffer in der Term-Datenbank gefunden werden können. Der Google Übersetzer und SDK BeGlobal Enterprise stehen kostenlos zur Verfügung und unterbreiten Vorschläge. Trados Studio trennt die zu übersetzenden Texte von der Codierung (Programmsprache) und präsentiert sie im Arbeitsfenster, damit man sich auf die Übersetzung konzentrieren kann.

Nur eine von den beiden im Modulhandbuch auftauchenden Grafiken braucht man zu lokalisieren. Das Logo der Hochschule Anhalt wird nicht lokalisiert. Für die Lokalisierung der Grafik wurde die Grafiksoftware Microsoft-Paint angewandt, da die Grafik nur einen weißen Hintergrund und ein paar Zeilen mit Texten besitzt.

5.4 Durchführung der Lokalisierung

In der Praxis beziehen sich die Lokalisierungsprozesse meist auf die folgenden Phasen: die Vorbereitung der Ausgangsmaterialien, die Vorbereitung der Referenzdateien, die Übersetzung der Texte, die Verwaltung der produktbezogenen Daten (TM, Terminologie usw.), die Anpassung der technischen Aspekte, die Qualitätssicherung und die Auslieferung der Endversion des Softwareproduktes.

5.4.1 Vorbereitung der zu lokalisierenden Quelldateien

Alle Vorbereitungen zielen auf eine reibungslose Durchführung der Lokalisierungsarbeit ab. Für eine CAT ist es erforderlich, die zu übersetzenden Dateien für das Lokalisierungsprogramm bereitzustellen. Wie schon erörtert, dass ein EPUB-Dokument eine Anwendung ist, kann das Lokalisierungsprogramm das EPUB-Dokument nicht als einfache Datei betrachten. Weil die Lokalisierungswerkzeuge das Dateiformat EPUB nicht erkennen, ist es nicht möglich, das EPUB-Dokument direkt in das Lokalisierungsprogramm zu importieren und zu bearbeiten. Eine EPUB-Publikation besteht aus vielen Dateien. Für die Analyse soll die Dateierweiterung des

EPUB-Dokuments von .epub in .zip umbenannt werden. Somit muss die ZIP-Datei mit einer Komprimierungssoftware entpackt und die in der ZIP-Datei liegenden Dateien analysiert sowie für das Importieren in der SDL Trados Studio vorbereitet werden. Für die Lokalisierung können die zu übersetzenden Dateien in einem anderen Ordner kopiert werden.

Außer der Inhaltsdateien und Grafiken ist noch das Inhaltsverzeichnis im Programm zu zeigen. Das heißt, die NCX- oder XHTML-Inhaltsverzeichnisdatei muss in Trados Studio importiert werden. Bei der Übersetzung der Strukturdateien dürfen die Verweise auf bestimmte Dateien oder Metadaten nicht geändert werden. Die Hyperlinks innerhalb des EPUB-Dokuments dürfen nicht beschädigt werden und müssen entsprechend in allen Dateien des EPUB-Dokuments übereinstimmen.

Dateien der folgenden Formate müssen im Rahmen des Projekts lokalisiert werden:

- *.jpg
- *.xhtml
- *.ncx

5.4.2 Konfigurieren des NCX-Dateityps in SDL Trados Studio

Das XHTML-Dateiformat wird von SDL Trados Studio unterstützt, während das NCX-Dateiformat nicht als Vorlage in Trados Studio besteht. Vor dem Importieren der NCX-Datei muss der Dateityp vom Benutzer definiert werden.

Ein neuer Dateityp kann in den Projekteinstellungen von Trados Studio mit Hilfe des Assistenten angelegt werden. Da eine NCX-Datei eigentlich eine XML-Datei ist, wird hier die Regelbasis von XML ausgewählt (Siehe Abbildung 14). Die Basisstruktur entspricht deshalb den Grundregeln von XML.

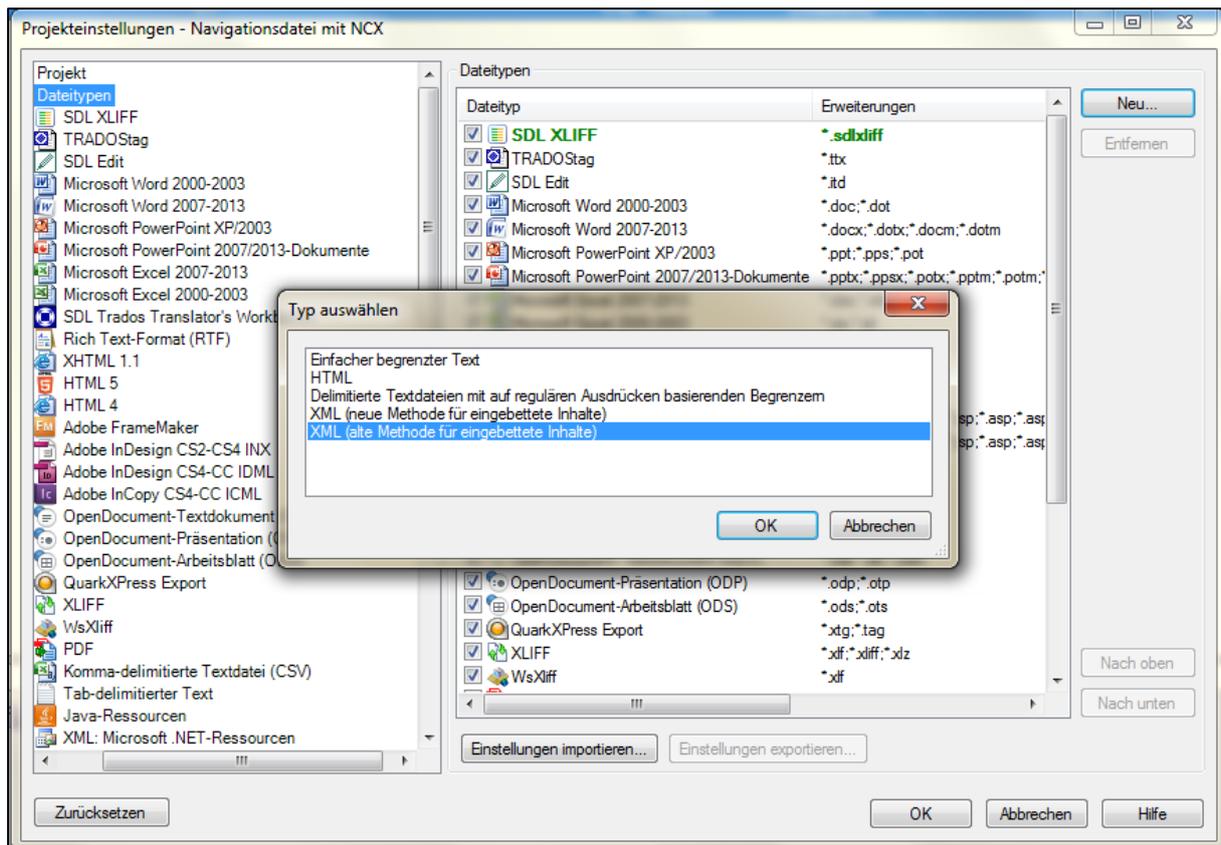


Abbildung 15. Auswahl der XML-Basis für NCX-Dateien

Die grundlegenden Informationen des Dateityps können im Dialogfeld Dateityp-Informationen ausgefüllt werden (Siehe Abbildung 15).

The screenshot shows a dialog box titled "Dateityp anlegen" (Create File Type) with a close button (X) in the top right corner. The main heading is "Dateityp-Informationen" (File Type Information) with a sub-instruction: "Geben Sie die Informationen an, die diesen Dateityp beschreiben." (Provide the information that describes this file type). A small icon of a document with a key is visible in the top right of the main area.

The dialog is organized into several sections:

- Dateityp-Informationen:**
 - Dateityp-Name:** A text field containing "NCX".
 - Dateityp-Symbol:** A button with a file icon and the text "assembly://Sdl.FileTypeSupport.Native.Xml_1_2/Sdl.FileTyp", followed by a "Durchsuchen..." (Browse...) button.
 - Dateityp-Kennung:** A text field containing "Kopie von XML v 1.2.0.0".
- Dateinamen (File Names):**
 - Name des einzelnen Dokuments:** A text field containing "NCX-Dokument".
 - Name der Dokumentkategorie:** A text field containing "NCX-Dokumente".
 - Dateinamenerweiterungen (Wildcards):** A text field containing "*.ncx".
- Beschreibung (Description):** A large text area containing the text "Navigation Control file for XML |".

At the bottom of the dialog, there are five buttons: "Hilfe" (Help), "< Zurück" (Back), "Weiter >" (Next), "Fertig stellen" (Finish), and "Abbrechen" (Cancel).

Abbildung 16. NCX-Informationen in SDL Trados Studio

Parser-Regeln beschreiben die Struktur eines Dateityps. Bei der Definition von Parser-Regeln ist ein Blick in den Quelltext der NCX-Datei zu empfehlen, um übersetzbaren vom nicht übersetzbaren Text zu trennen. Zugleich sollen die Struktur-Tags von Inline-Tags unterschieden werden. Dadurch wird herausgefunden, dass die zu übersetzenden Zeichenketten innerhalb der Anfang- und End-Elemente <docTitel>, <docAuthor>, <navLabel> usw. stehen und sich im Inline-Tag <text> befinden (siehe Abbildungen 16 und 17).

```
22     <meta name="dtb:maxPageNumber" content="0"/>
23   </head>
24
25
26 <docTitle><text> Modulhandbuch </text></docTitle>
27 <docAuthor><text> Softwarelokalisierung </text></docAuthor>
28
29 <navMap>
30 <navPoint id="deck" playOrder="1">
31 <navLabel><text>Deckblatt</text></navLabel>
32 <content src="deckblatt.xhtml"/>
33 </navPoint>
34 <navPoint id="modul01" playOrder="2">
35 <navLabel><text>Effizientes Terminologiemanagement</text></navLabel>
36 <content src="modul01.xhtml"/>
37 </navPoint>
38 <navPoint id="modul02" playOrder="3">
39 <navLabel><text>IT- und translationswissenschaftliche Grundlagen der
40 Softwarelokalisierung</text></navLabel>
41 <content src="modul02.xhtml"/>
42 </navPoint>
43 <navPoint id="modul03" playOrder="4">
44 <navLabel><text>Lokalisierung grafischer Benutzungsoberflächen</text>
45 <content src="modul03.xhtml"/>
46 </navPoint>
47 <navPoint id="modul04" playOrder="5">
48 <navLabel><text>Übersetzungs- Lokalisierungstechnologie
```

Abbildung 17. Analyse des NCX-Quelltextes

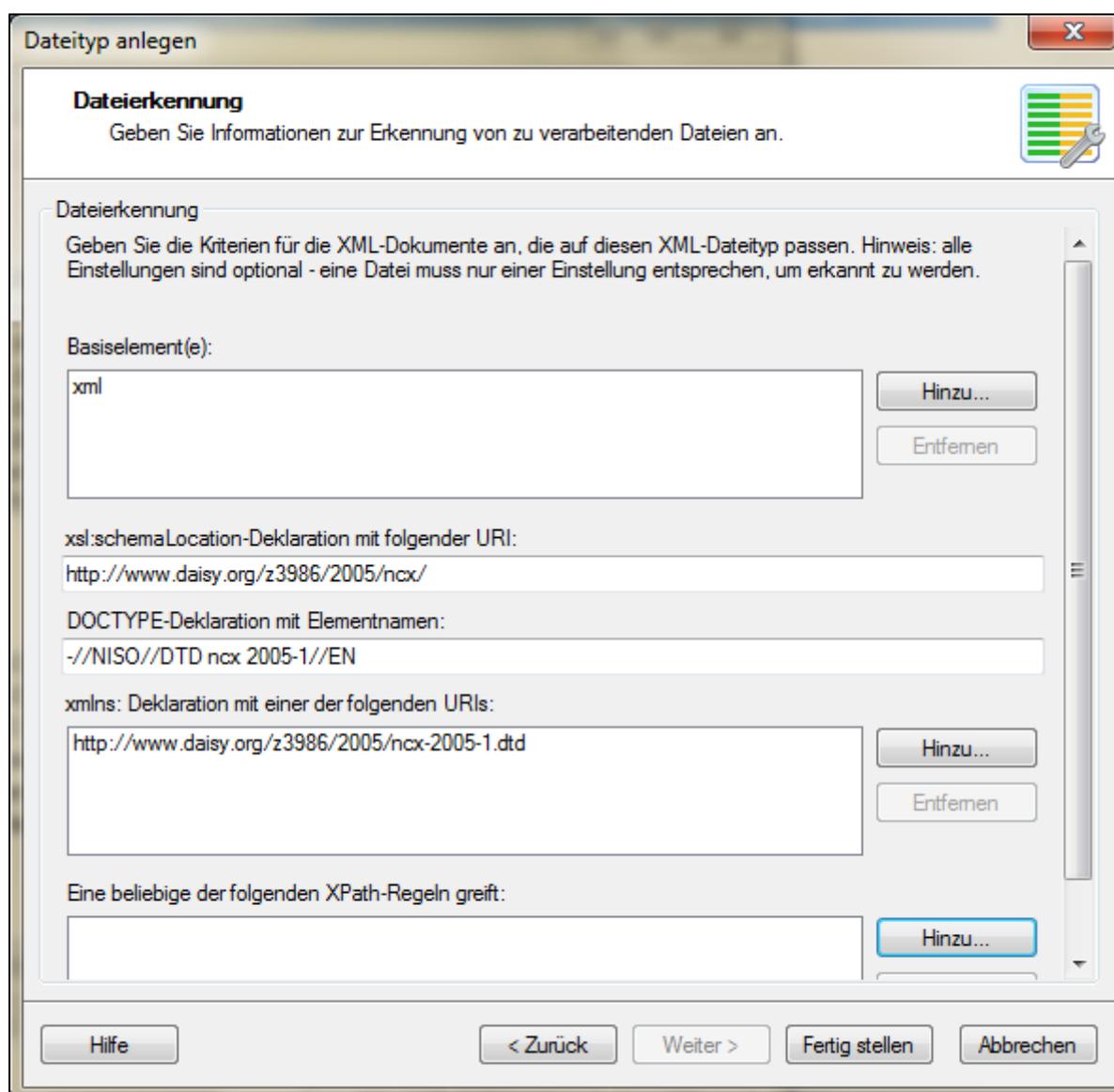


Abbildung 18. Dateierkennung von NCX in Projekteinstellungen

5.4.3 Vorbereitung der Referenzdateien

Eine erfolgreiche Lokalisierung setzt eine eindeutige Terminologie voraus. Deshalb soll eine Termbank (Terminologiedatenbank) vor der Lokalisierung vorbereitet werden. Diese Termbank dient während des Lokalisierungsablaufs einerseits als Wörterbuch und wird andererseits für die Qualitätssicherung der Terminologiekonsistenz eingesetzt, sodass die Übersetzung eines neuen Eintrags mit den bereits gespeicherten Übersetzungen verglichen werden kann.

Die Terminologie lässt sich mithilfe von Terminologieextraktionsprogrammen gewinnen. SDL MultiTerm Extract wurde speziell für Übersetzer und Terminologen entwickelt. Das Softwareprodukt bietet mithilfe moderner Extrahierungsmethoden eine ein- oder mehrsprachige Terminologieerkennung an. Quelldateien von unterstützten Formaten können durch verschiedene Termlängen gegliedert werden. Doch ist nichts in der Welt perfekt. Ein vollkommenes Ergebnis ist nicht einfach zu bekommen, weshalb eine andere Methode in diesem Projekt vorgezogen wurde.

Im vorliegenden Projekt gibt es nicht viele Unterlagen und Informationen. Eine manuelle Extraktion mithilfe von Microsoft-Excel reicht in diesem Fall für die Terminologiearbeit aus. Mittels der Funktionen von Word und Excel können Termini extrahiert werden, z.B. mithilfe von Makros und einer Stopwortliste⁸⁵. Es handelt sich bei dem EPUB-Dokument um 16 Modulbeschreibungen. Die einzelnen Dateien belegen die gleiche Inhaltsstruktur mit gleichen Wörtern. Diese häufig auftauchenden Wörter können als Termini in Excel gesammelt werden. Die entsprechenden chinesischen Übersetzungen werden auch hinzugefügt und die Eigenschaften der Termini zusätzlich ausgefüllt. Mit der Excel-Datei wird das Glossar des Projekts an erster Stelle zusammengefasst. Mithilfe der SDL Convert und SDL MultiTerm wurde die initialisierte Termbank erstellt. Beim Erstellen des Projekts wurde die Termbank als ein Referenzmaterial hinzugefügt.

Als TBX-Datei ist die Termbank für die weitere Arbeit in SDL Trados Studio vorbereitet. Die Termbank ist bei der Einstellung für die Bearbeitung aktiviert. Damit können die bei der Übersetzung zutreffenden Termini auch hinzugefügt werden (Siehe Abbildung 19). Die durch roten Rahmen hervorgehobenen Felder beziehen sich auf die Termbank. Die Terminologiearbeit wie Nachschlagen kann im integrierten Programm leicht zum Einsatz kommen.

⁸⁵ Eine Stopwortliste ist eine Sammlung der allgemeinen üblichen Stopwörter. Stopwörter sind die Wörter, die bei einer Volltextindexierung nicht beachtet werden, da sie sehr häufig auftreten und gewöhnlich keine Relevanz für die Erfassung des Dokumentinhalts besitzen, z.B. die bestimmten Artikel ('der', 'die', 'das'), die unbestimmten Artikel ('einer', 'eine', 'ein'), Konjunktionen (z. B. 'und', 'oder', 'doch'), gebrauchte Präpositionen usw. (Vgl. Stopwort (2015): Stopwort. Online im Internet: <https://de.wikipedia.org/wiki/Stopwort> [09.09.2015])

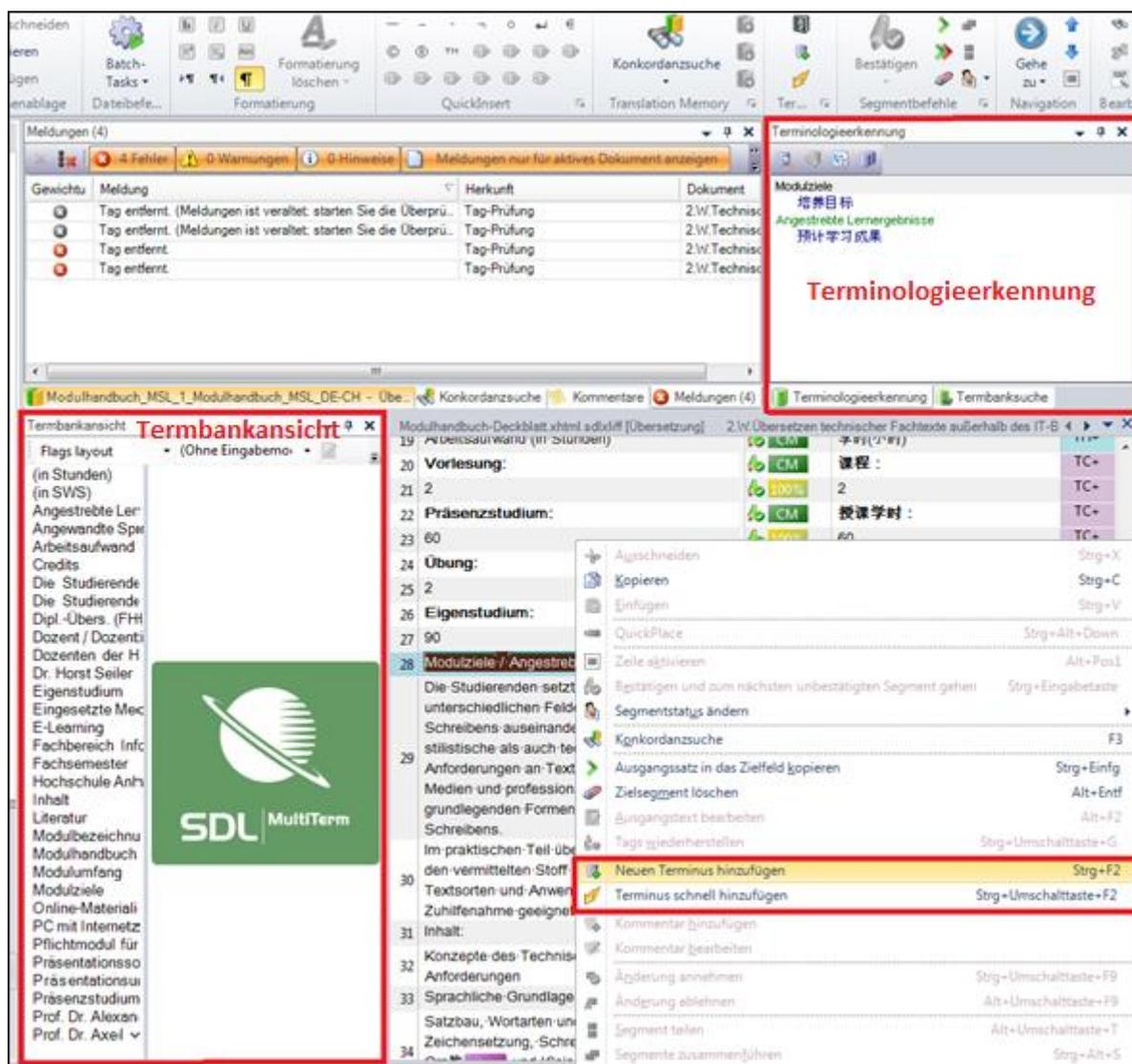


Abbildung 19. Hinzufügen neuer Terminologie in SDL Trados Studio 2014

Weil das Projekt ein einzelnes Lokalisierungsprojekt ist, gibt es kein vorhandenes TM (Translation Memory), demnach wurde ein neues TM beim Erstellen des Projektes angelegt und hinzugefügt. Das TM-System speichert alle Segmentpaare während der Übersetzung in der Datei *Modulhandbuch_MSL_DE-CH.sdlzm* ab. Für die weitere Arbeit im relevanten Gebiet kann man die erzeugte Terminologiedatenbank sowie das TM als Referenzmaterialien verwenden. Je mehr mehrstufige Segmentpaare die beiden Referenzdateien einschließen, desto eindeutiger ist der übersetzte Text. Das heißt, wenn eine Wortgruppe im TM eine Übersetzung hat, dann ist es hilfreich, dass auch die einzelnen Wörter in der Wortgruppe eine eigene Übersetzung besitzen. Damit werden die ähnlichen Ausgangssegmente genau in der Zielsprache übertragen und gleich gehalten.

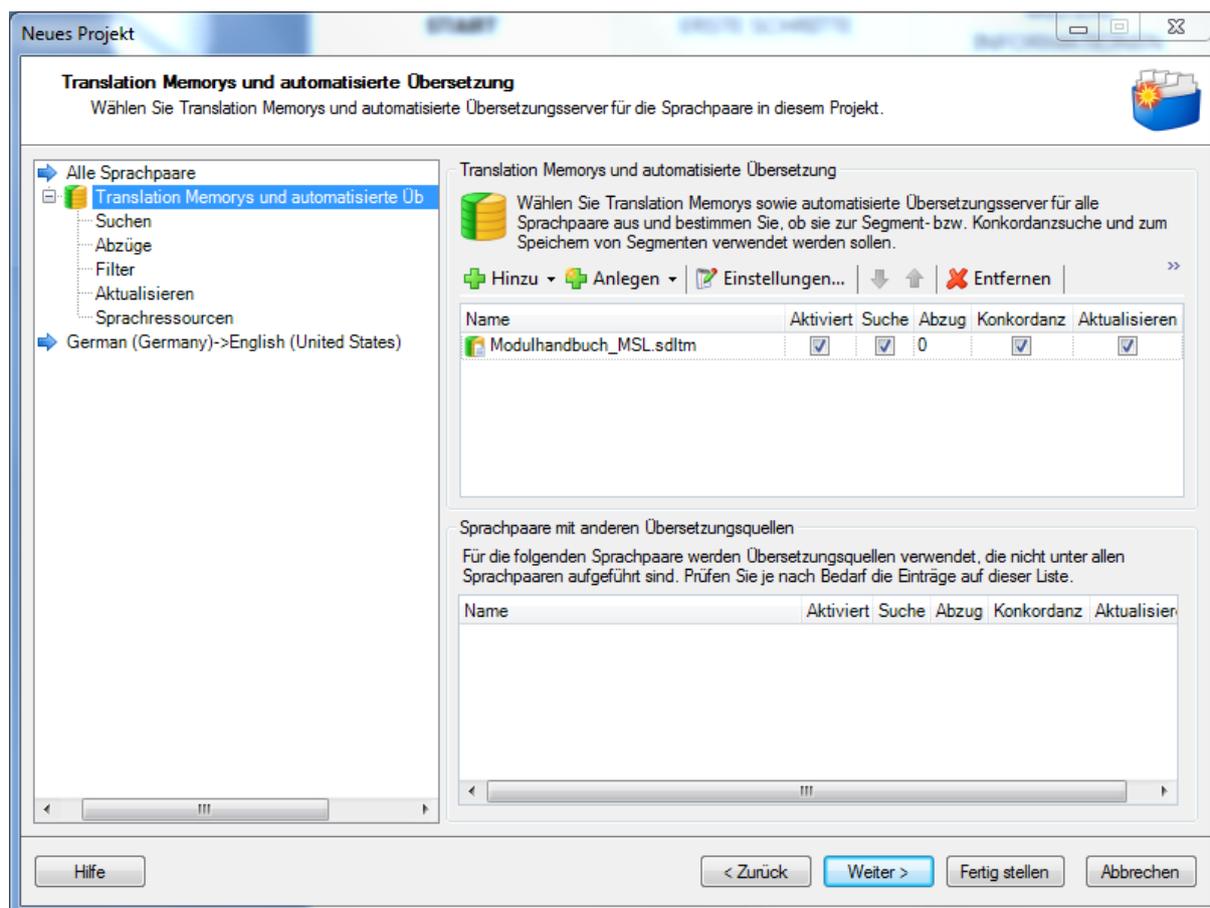


Abbildung 20. Anlegen eines Translation Memrys in SDL Trados Studio 2014

5.4.4 Übersetzung

Auf inhaltlicher Ebene müssen Änderungen mit den im Lande üblichen Konventionen vorgenommen werden. Die zu übersetzenden Wörter dürfen nicht reduziert oder erweitert werden. Die Formulierungen müssen an die chinesischen Gewohnheiten angepasst werden.

Die Zielgruppe des lokalisierten Dokuments sind die chinesischen Studierenden im Studiengang Master-Softwarelokalisierung sowie die sich für das Studium interessierenden Chinesen. Der staatliche Standard des Referenzdokumentationsformats (Chinesisch: 参考文献格式国家标准) empfiehlt das Literaturverzeichnis am Ende einer Literatur grundsätzlich mit derselben Sprache wie die Literatur selbst [参考文献格式国家标准: 5.1.1]. Die Groß- und Kleinschreibung

der Buchstaben muss mit der Redewendung der Ausgangssprache im Einklang stehen [参考文献格式国家标准: 5.3]. Das Datum soll mit arabischen Ziffern dokumentiert werden und das Datumsformat lautet JJJJ-MM-TT[参考文献格式国家标准: 6.7.3.1].

Im zweiten Fall ist diese Gruppe von Interessierten mit den fachlichen Begriffen nicht vollkommen vertraut und ihr Wissensstand anders als Menschen in Deutschland, daher ist eine Erklärung der Abkürzung für die Einsteiger hilfreich, wie zum Beispiel „LISA“ als die Abkürzung von „Localization Industry Standards Association“ (Siehe Abbildung 21). Auf diese Weise werden den Interessierenden mehr eindeutige Informationen angeboten. Dadurch können Verwechslungen vermieden werden.

48	LISA-QA-Model-sowie-QS-Metriken-/Modelle	 LISA (Localization-Industry-Standards-Association) QA (Qualitätssicherung) 模式和QS度量/模式	L+
----	--	--	----

Abbildung 21. Problematik der Abkürzungen in der Übersetzung

Besonders bei der Übersetzung von Personen- und Eigennamen wird keine dominierende Meinung im Übersetzungsmarkt in China als Standard festgelegt. Bei der Übersetzung von Personen- und Eigennamen ist zu bestimmen, ob es in China entsprechende offizielle chinesische Übersetzungen von Namen gibt, die die Öffentlichkeit schon zur Kenntnis genommen hat. Das öffentliche Bewusstsein und die offizielle Veröffentlichung spielen bei der Bestimmung der Darstellung im Zieltext eine wesentliche Rolle. Wenn es keine offizielle Übersetzung gibt und darüber hinaus eine eigene Übersetzung Missverständnis verursacht und damit sich nicht auf die richtigen Quellen beziehen kann, soll der Name in der originalen Form beibehalten werden. Um einen vernünftigen Zieltext bieten zu können, ist es notwendig am Anfang den Namen oder den Begriff in der Zielsprache zu recherchieren.

47	Literatur:	 100%	课程相关参考书目 :
48	Duden.		杜登德语词典
49	Die-deutsche-Rechtschreibung.		德语正字法
	<u>25.</u> ,-völlig-neu-bearbeitete-und-erweiterte-Aufl.-(=Duden,-Bd.1)		25.,-völlig-neu-bearbeitete-und-erweiterte-Aufl.-(=

Abbildung 22. Ins Chinesischen übertragende Literaturnamen

Für neue Begriffe kann eine Ausdrucksweise durch lange Diskussion und die Wahl des Volkes erst angenommen werden. Solche neuen Übersetzungen hängen nicht nur von den sprachlichen Erkenntnissen und Kompetenz der Übersetzer, sondern auch von ihrer Kreativität und ihrem Einfluss ab. Wenn die bibliografischen Angaben in die Zielsprache übersetzt werden müssen, kann eine Angabe des entsprechenden Originaltitels in Klammern angeboten werden. Die Ludwig-Maximilians-Universität in München hat auch eine Ausnahme beim Zitieren erhoben: „denkbar ist dies etwa bei theoretischen Texten, die keinen zentralen Status für die Arbeit besitzen und bei der es nicht auf den Wortlaut ankommt; weitere Ausnahmen sind mit den wichtigsten intendierten Lesern der entsprechenden Arbeit abzusprechen“⁸⁶.

Obwohl der Name der originalen Form im Chinesischen beibehalten wird, muss das Nomen an die Zielkultur angepasst und umgeformt werden. Zum Beispiel gibt es im Chinesischen normalerweise keine Pluralformen von Nomen. Die Pluralform von chinesischen Nomen können meistens nicht durch Flexionen, sondern hauptsächlich durch eine Kombination von syntaktischen Elementen erreicht werden, um die Pluralbedeutung auszudrücken. Die Deklinationen werden im chinesischen Kontext einfach entfernt (Siehe Abbildung 23). Weiterhin verwendet man im chinesischen Schriftstil keine Leerzeichen. Nur wenn mehr als zwei Wörter aus einer indogermanischen Sprache u.ä. nebeneinander stehen, wird ein Leerzeichen zwischen den Wörtern sowie ihren Satzzeichen benötigt.

⁸⁶ Vgl. Ludwig-Maximilians-Universität München (o.J.): Hinweise zu Zitiertechniken und bibliographischen Angaben.

41	DTDs, Schemata, Stylesheets-etc.		DTD, Schema, Stylesheet等	LI+
----	----------------------------------	---	--------------------------	-----

Abbildung 23. Eigennamen in den originalen Formen (Nur das Wort etc. ins Chinesische übersetzt)

5.5 Test der lokalisierten Anwendung

Um ein überzeugendes Ergebnis zu erzielen, müssen zwei Schritte durchgeführt werden: Die inhaltliche Überprüfung und der Verfügbarkeitstest der Anwendung. Die inhaltliche Überprüfung kann durch maschinelle Funktionen und menschliches Korrekturlesen durchgeführt werden. Für den Verfügbarkeitstest müssen Leseprogramme für E-Books zum Einsatz kommen. Falls beim Anzeigen des EPUB-Dokuments Probleme auftreten, müssen die Materialien manuell auf technische Fehler überprüft werden.

5.5.1 Qualitätssicherung

Das Überprüfungssystem von SDL Trados Studio kann durch Einstellen verschiedener Optionen konfiguriert werden. Die softwaregestützte Rechtschreibprüfung sucht die Rechtschreib- und Tippfehler, die zu orthografischen, grammatikalischen und semantischen Fehlern gehören können. Deutsch und Chinesisch stammen aus verschiedenen Sprachfamilien. Der Satzbau und die Ausdrucksweise beider Sprachen sind völlig anders. Bei der QA-Überprüfung von chinesischen Texten liefern die elektronischen Programme relativ schlechte Ergebnisse. Für Chinesisch ist die Überprüfung von Leerzeichen und der Reihenfolge der Wörter nicht sinnvoll. Wichtig ist hingegen die Überprüfung der Terminologiekonsistenz. Auf der semantischen Ebene erfordert es nicht nur die Einhaltung der Genauigkeit der Übertragung, sondern auch die Vollständigkeit des Inhalts. Für die äquivalente Übertragung der Bedeutung reicht die automatische Überprüfung des Programms nicht aus. Ein muttersprachliches Lektorat ist sowohl

für die sprachlichen und inhaltlichen Aspekte als auch für die stilistischen Aspekte erforderlich. Hier empfiehlt sich der Einsatz des Vier-Augen-Prinzips.

Auch die Ausgangsmaterialien können Rechtschreib- und Formatfehler enthalten und sogar mehrdeutig bzw. inkonsistent sein. Während der Lokalisierung mit einem Lokalisierungsprogramm wurden einige Inkonsistenzen der Ausgangstexte mit Hilfe eines TMs gekennzeichnet, wie zum Beispiel Großbuchstaben bzw. Kleinbuchstaben desselben Terminus, die durch menschliches Lesen schwer zu finden sind (Siehe Abbildung 24). In diesem Fall ist eine Änderung des Ausgangstexts notwendig.

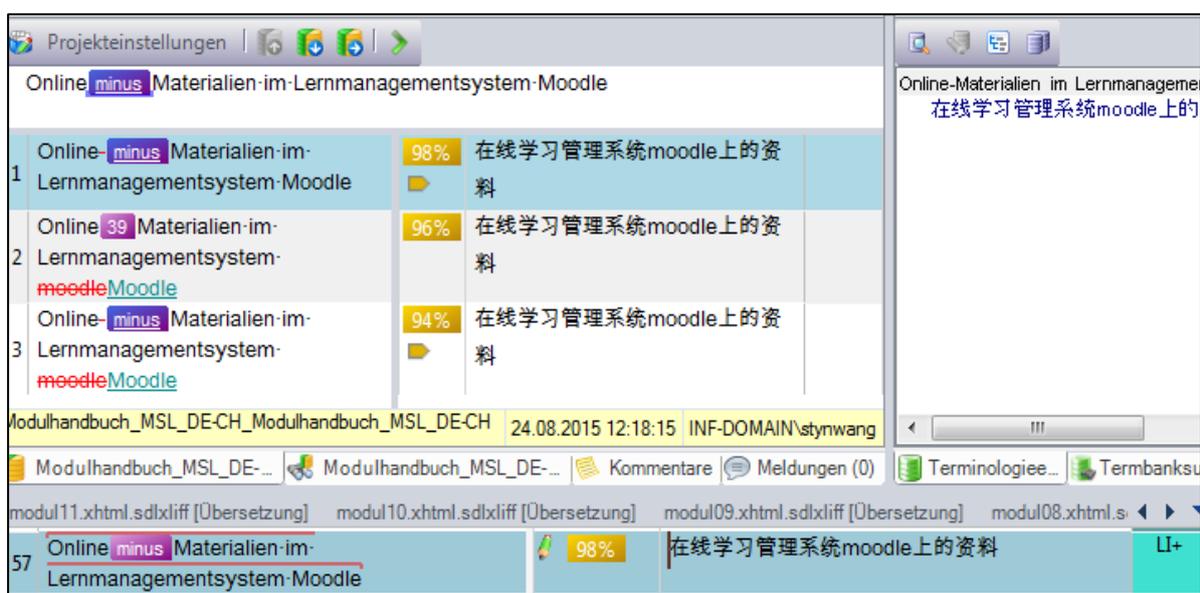


Abbildung 24. Mithilfe eines TMs gefundene Inkonsistenz im Ausgangstext

5.5.2 Erstellen des zielsprachlichen EPUB-Dokuments

Nach der Bestätigung aller durch die Qualitätssicherung überprüften Übersetzungen, werden die zielsprachlichen Dateien exportiert. Trados Studio generiert die Übersetzungen im selben Format, in dem die Ausgangstexte vorliegen. Das Erzeugen der zielsprachlichen EPUB-Version des Modulhandbuchs passiert erst nach dem Export aller lokalisierten Einzeldateien.

Die Inhaltsdateien, Vorlagendateien sowie die Grafiken müssen entsprechend der deutschen Version im OPS-Verzeichnis (Inhaltsdateien können eventuell in einem weiteren Unterverzeichnis durch den Ersteller oder ein bestimmtes Programm erstellt werden, wie das Unterverzeichnis *Text* von Sigil⁸⁷) bzw. im Unterverzeichnis *image* liegen. Ein chinesischer Name kann beim Erstellen der ZIP-Datei angegeben werden. Aufgrund von Zugriffsproblemen beim deutschen Windowssystem wurde kein chinesischer Name angegeben, sondern lediglich die Zusatzinformation *_ch* am Ende des Dateinamens eingefügt. Danach werden die *MIME-Type-Datei* unkomprimiert und die Verzeichnisse *META-INF* und *OPS* zur ZIP-Datei hinzugefügt. Schließlich soll die Namensweiterung des ZIP-Archivs durch *.epub* ersetzt werden.

Weiterhin braucht man für spezielle chinesische Schriftarten eine TTF-Datei. TTF steht für TrueType- Schriften und stellt einen Standard für Fonts zur Schriftdarstellung auf Bildschirmen und Druckern dar. In einer TTF-Datei können mehrere Schriften enthalten sein. TrueType-Schriften werden nach dem Prinzip einer Vektorgrafik aus Konturen aufgebaut. Der Vorteil der Vektordarstellung ist die verlustfreie Skalierbarkeit, d.h. die Größe der Schrift ist beliebig veränderbar.⁸⁸

5.5.3 Verfügbarkeitstest des chinesischen EPUB-Dokuments

Der Verfügbarkeitstest der Anwendung kann erst nach dem Generieren des chinesischen EPUB-Dokuments erfolgen. Die Prüfung erfolgt wie im Abschnitt 3.6 "Überprüfung von EPUB-Dokumenten" erläutert.

Es kann verschiedene Ursachen geben, dass ein lokalisiertes EPUB-Dokument nicht dargestellt werden kann. Beispielsweise kann die Unterstützung der Schriftart bei der verwendeten Darstellungssoftware fehlen. Außerdem ist es möglich, dass der Quellcode bei der Lokalisierung beschädigt wurde. Abschließend ist noch eine Überprüfung der eingebetteten Grafik vonnöten. Wurden alle Fehler beseitigt, wird

⁸⁷ Siehe

⁸⁸ Vgl. TrueType (2015): TrueType. Online im Internet: <https://de.wikipedia.org/wiki/TrueType> [09.09.2015]

das Dokument wie gewünscht dargestellt und kann ausgeliefert und verwendet werden.

6. Fazit und Ausblick

Die in dieser Arbeit an einem Beispiel durchlaufenen Phasen der Erstellung und Lokalisierung eines EPUB-Dokuments wurden vollständig durchgeführt. Die aufgetretenen Probleme wurden innerhalb des Projekts bewältigt.

Es wurde deutlich, dass zur selbstständigen Erzeugung eines EPUB-Dokuments zumindest Grundkenntnisse von (X)HTML und XML erforderlich sind. Des Weiteren ist es wichtig, dass sämtliche Dateien, die in einer Publikation dazu dienen, dass das Dokument richtig dargestellt werden kann - wie die Datei des Inhaltsverzeichnisses und die Dateien im OPF-Ordner - vorhanden sind. Die Zusammenführung aller Teildateien muss in der Form erfolgen, dass alle IDs und Verweise an jeglichen Stellen übereinstimmen. Weiterhin muss der einmalig angegebene Name oder die einmalig angegebene ID im gesamten Projekt beibehalten werden. Zugleich ist eine sorgfältige Überprüfung die Voraussetzung zur Ausführung eines erfolgreichen EPUB-Dokuments, eine Voraussetzung, die auch bei der Erstellung und Lokalisierung von Dokumenten anderer Formate zutreffend ist. Die Terminologieverwaltungs- und TM-Systeme sind eine gute Hilfe für die Lokalisierung von EPUB-Dokumenten.

Unter Berücksichtigung der kulturellen Besonderheiten muss bei der Erstellung und Lokalisierung jeweils auf die Internationalisierung und die kulturellen Besonderheiten geachtet werden, sodass die Arbeitsprozesse reibungslos durchgeführt und Kulturbarrieren vermieden werden können.

In der Praxis lassen sich Bücher mit komplexen Formaten nur teilweise von E-Book-Readern interpretieren. Ein fehlerloses EPUB-Dokument erfordert nicht nur gut strukturierte Teildateien, sondern auch angepasste Hardware und Software für das Anzeigen. In der Zukunft können die noch vorhandenen Probleme aber wohl durch die Zusammenarbeit von „Amateuren“ mit der IDPF-Organisation gelöst und das Format weiterentwickelt werden. Damit kann dem Leser dann ein noch besseres Leseerlebnis geboten werden.

Literaturverzeichnis:

[A3W 2012] A3W. E-Book. Online im Internet. <<http://www.a3w.de/e-book-2/>>, 2015.09.09

[Broda 2013] S. Broda. Technische Dokumentation auf mobilen Endgeräten. Online im Internet. <<https://www.texniq.de/de/blog/mobile-dokumentation>>, 2015.09.09

[Data2type o.J.] Data2type. ePUB. Online im Internet. <<http://www.data2type.de/xml-xslt-xslfo/epub-ueberblick/>>, 2015.09.09

[Deutschbein o.J.] R. Deutschbein. E-Book-Reader als Alternative zum gedruckten Buch. Online im Internet. <<http://www.teltarif.de/h/e-book-reader.html>>, 2015.09.09

[ebooknet 2014] ebooknet. Was sind eBooks? Online im Internet. <<http://www.ebooknet.de/2014/was-sind-ebooks>>, 2015.09.09

[em 2014] em(Schriftsatz). wikipedia. Online im Internet. <[https://de.wikipedia.org/wiki/Em_\(Schriftsatz\)](https://de.wikipedia.org/wiki/Em_(Schriftsatz))>, 2015.09.09

[FontShop 2015] FontShop. Was ist OpenType? <<http://www.fontshop.de/wissen/ist-opentype/>>, 2015.09.09

[Gardiner 2010] E. Gardiner, R. G. Musto. The electronic book. Chapter 19. Online im Internet. <<http://www.wsj.com/articles/SB10001424052748704187204575102110426333220>>, 2015.09.09

[Handner 2011] K. Handner. ePUB für die Technische Dokumentation? Online im Internet. <<http://www.klaretex.de/magazin/artikel/epub-fur-die-technische-dokumentation/>>, 2015.09.09

[Hauschild 2014] J. Hauschild. Neue Lesegewohnheiten: In F-Form durch Texte springen. Online im Internet. <<http://www.spiegel.de/wissenschaft/mensch/lesen-im-internet-veraenderungen-der-gewohnheiten-a-971179.html>>, 2015.09.09

[IDPF 2015] International Digital Publishing Forum < <http://idpf.org/about-us> >, 2015.09.09

[InDesign 2015] Adobe InDesign. Online im Internet. <https://de.wikipedia.org/wiki/Adobe_InDesign>, 2015.09.09

[ITWissen 2015] ITWissen. OeBF (open ebook forum). Online im Internet. <<http://www.itwissen.info/definition/lexikon/open-ebook-forum-OeBF.html>>, 2015.09.09

[Kirch 2014] D. Kirch. Zu Industrie 4.0 durch digitale technische Dokumentation. Online im Internet. <<https://www.de.capgemini-consulting.com/blog/digital-transformation-blog/2014/07/eintrag-digitale-technische-dokumentation>>, 2015.09.09

[Low 2011] T. Low. Verlagsgeplauder - Ebooks – technische Revolution oder teure Sackgasse? Online im Internet. <<http://verlorenerwerke.blogspot.de/2011/07/verlagsgeplauder-ebooks-technische.html>>, 2015.09.09

[München o.J.] Ludwig-Maximilians-Universität München. Hinweise zu Zitiertechniken und bibliographischen Angaben.

[Otte 2009] Tobias Otte. Bindestrich in HTML. Online im Internet. <<http://tobias-otte.de/2007/05/bindestrich-in-html/>>, 2015.09.09

[selfhtml 2015] selfhtml. Zeichencodierung. Online im Internet. <<http://wiki.selfhtml.org/wiki/Zeichencodierung#ASCII>>, 2015.09.09

[Semcon 2013] Semcon Communication GmbH. Dokumentation auf mobilen Endgeräten - Neuauflage für das Single Sourcing. Online im Internet. <http://www.doku.info/doku_article_447.html>, 2015.09.09

[Siegel 2010] S. Siegel. Technische Dokumentation auf E-Book Readern – geht das? Online im Internet.

<http://www.tekom.de/uploads/media/208/POD4_Siegel_20813.pdf>, 2015.09.09

[Singh 2005] N. Singh, A. Pereira. The culturally customized web site. Customizing web sites for the global marketplace. 1. Auflage. Elsevier Butterworth-Heinemann. 2005

[Sprachenquilt 2009] sprachenquilt. Gedankenstrich und Bindestrich, “en dash” und “em dash” und die Leertaste. Online im Internet.

<<http://sprachenquilt.com/2009/03/26/gedankenstrich-und-bindestrich-en-dash-und-em-dash-und-die-leertaste/>>, 2015.09.09

[Transcom o.J.] Transcom. Was ist "Technische Dokumentation"? Online im Internet.

<<http://www.transcom.de/transcom/de/wirueberuns/leistungen/technische-dokumentation.htm>>, 2015.09.09

[Warner 2012] A. Warner. Vom Buch zum Byte. Online im Internet.

<<http://www.netzpiloten.de/vom-buch-zum-byte-die-geschichte-der-e-books/>>, 2015.09.09

Chinesische Quellen:

[Niu 2012] Niu. EPubSTAR. Online im Internet. <http://www.by-smart.com/eread/eread_11.html>, 2015.09.09

[全国工作标准化技术委员会 o.J.] 全国工作标准化技术委员会. 参考文献格式国家标准. (Selbst Übersetzung) Kommission des nationalen technischen Standards. Staatlicher Standard der Literaturangaben.

Anhang

Anhang A: Modulhandbuch_MSL.sdlfb

Anhang B: Modulhandbuch_MSL_DE-CH.sdlfb

Anhang C: Modulhandbuch_MSL_DE_Version2.epub

Anhang D: Modulhandbuch_MSL_DE_Version3.epub

Anhang E: Modulhandbuch_MSL_ZH_Version2.epub

Anhang F: Modulhandbuch_MSL_ZH_Version3.epub

Anhang G: Modulhandbuch_MSL_Sigil.epub

Anhang H: Modulhandbuch_MSL_ZH_Sigil.epub

Eidesstattliche Erklärung

„Diese Arbeit wurde von mir selbständig verfasst und in gleicher oder ähnlicher Fassung noch nicht in einem anderen Studiengang als Prüfungsleistung vorgelegt. Ich habe keine anderen als die angegebenen Hilfsmittel und Quellen, einschließlich der angegebenen oder beschriebenen Software, verwendet.“

Datum: _____

Unterschrift: _____