

Volltexte im Internet

Einleitung

- werde sprechen über:
 - Definitionen von Volltext
 - Gewinnung von Volltext
 - Wo uns Volltext begegnet
 - Welche Punkte des Urheberrechts beachtet werden müssen.
- Wort Volltext häufig gebraucht
- *Was fällt euch zu Volltext ein?* → Anfertigen eines Stichwortkataloges

Definitionen

- weitere und engere → gehe auf die engeren ein → weitere zu weitführend
- **Folie auflegen**

Full-text refers to unstructured, free, natural language text. This term is usually used in contrast to fixed-structure, record-oriented, or otherwise restricted text (...).¹

Volltext ist der gängige Fachbegriff für den eigentlichen Textkorporus – incl. Überschriften, Fußnoten u.ä. – einer Publikation Im Gegensatz zum Freitext sind damit also nicht die typischerweise von Bibliothekaren oder Dokumentaren beigefügten Sekundärdaten zur formalen und inhaltlichen Erschließung gemeint.²

Was macht Volltext aus?

- reiner Textkorporus
- mit Überschriften und Fußnoten → die jedoch nicht mehr als solche gekennzeichnet
- u.U. Umlaute aufgelöst (ä in ae, ö in oe, ü in ue), nur einige Sonderzeichen darstellbar
- reiner Text ohne Strukturierung und Bilder
- Bilder nur, wenn sie sich mit OCR in Text umwandeln lassen (genauerer Eingehen auf OCR später)
- ASCII

Wie wird Volltext gewonnen?

- a) viele Texte heute bereits in irgendeiner Form elektronisch vorhanden (z.B. als HTML)
→ Auszeichnungen werden gelöscht (bei HTML z.B. alles was zwischen <> steht)
- o. Text einfach in txt-Format als reiner ASCII-Text gespeichert

- b) Texte werden retrospektiv digitalisiert
→ Einscannen – Images
+ OCR-Erkennung (Optical Character Recognition) → Umwandlung in ASCII
oder: manuelles Erfassen

- historische Schrifttypen + handschriftliche Bücher besonderes Problem, weil maschinelles Lesen (OCR) wenig erfolgreich und immer weniger Menschen Beherrschung dieser Schrift zu 100%
- weiteres Problem Bilder
→ meistens nur Faksimilierung
- in einigen Fällen Hinterlegung Faksimilierung mit Volltext
- häufig parallel zum Faksimile Text im Volltext, z.B. <http://www.gutenbergdigital.de>

¹ Full-Text and Information Retrieval, <http://www.iath.virginia.edu/elab/hfl0112.html>, Stand: 04.06.02

² Achim Oßwald, Elektronische Volltexte im Internet, http://www.lub.lu.se/UB2proj/LIS_collection/osswald-bt95/bt95-3.html, Stand: 03.06.02

Wo begegnet uns Volltext?

- normaler Editor auf dem PC
- Dateiende txt – reiner ASCII-Code

Beispiel für Volltext im Internet: <http://www.history-journals.de/hjn-04-01.txt>

- ♦ Volltextdatenbank
- ♦ Volltextsuche, Volltextretrieval, Volltextrecherche
- ♦ Volltextspeicherung
- ♦ Volltextverarbeitung

Beispiele für Volltext im Internet:

1) Google

- bei Erarbeitung des Themas → Suche nach Definitionen
- „full text“ AND „library terminology“ OR Volltext AND Bibliotheksterminologie

bei Google auch suchbar PDF-, HTML-, DOC-Dokumente

→ Dokumente auf reinen Text zurückgeführt

→ diesen Volltext Wort für Wort indexiert

→ Stoppwörter entfernen (z.B. und, die, .de)

→ Wörter in Feldern abgelegt

geschieht alles automatisch

→ durchsuchbar gemacht

Bsp. PDF-Dokumente als HTML-Version anzeigbar

Suchmöglichkeiten der Volltextsuche: Bool'sche Operatoren, Phrasensuche i.d.R. möglich

2) Gutenberg Projekt USA

- <http://www.gutenberg.net/>
- Suchmaschine: <http://promo.net/cgi-promo/pg>
- Eingabe Autor: Jane Austen
- Ausgabe Letters To Dead Authors (Jane Austen kommt im Volltext vor)

3) German Leeds

- Suchmaschine: <http://www.german.leeds.ac.uk/textarch1.htm>
- zuerst Auswahl des zu durchsuchenden Textes, z.B. Effie Briest
- danach Beginn Suche: z.B. Ostsee
- Fundstelle wird kurz angezeigt, Wort als Link ausgezeichnet
→ Verlinkung auf ein HTML-Dokument, in das der Text zwischen die Tags `<pre></pre>` aus der Datenbank eingespeist wird

Angebote im Internet, die auf „Veredlungen“ von Volltexten beruhen

1) GERHARD

- Volltextsuche
- Volltextdatenbank

→ automatische Indexierung (Rückführung Text auf Volltext, Format ASCII)

+ automatische Klassifizierung nach der UDK

→ Texte werden anhand der enthaltenen Worte klassifiziert, d.h. mit Sekundärdaten versehen

→ Notationen (die gleich in Begriff verfasst)

durchführen einer kurzen Suche bei GERHARD

2) Science Citation Index

- Literaturverzeichnisse im Volltext
- Suche nicht möglich, da kommerzieller Anbieter

3) JURIS-Datenbank

- Online-Volltextsuche möglich
- Vorführung nicht möglich, da kommerzieller Anbieter und ich keinen Zugriff von hier

→ Lizenzsache

4) Endprodukt

a) Printprodukt (Printing on demand)

der Volltext liegt elektronisch vor und wird entsprechend ausgezeichnet (z.B. als PDF-Dokument) wird auf Bestellung ausgedruckt und als Buch gebunden (gut geeignet für kleine Auflagen oder für bisher als vergriffen gegoltene Bücher)

b) Elektronisches Buch (eBook)

der Volltext liegt elektronisch vor und wird entsprechend strukturiert (z.B. als HTML-, PDF- oder XML-Dokument)

das Buch kann heruntergeladen oder als Datei zugeschickt werden
zum Lesen ein Reader benötigt

Urheberrecht

Grundlage für digitalisierte Texte immer Texte eines Urhebers

→ damit wichtig Beachtung Urheberrecht

- kurz nochmals ins Gedächtnis rufen, warum Digitalisierung notwendig
- säurehaltiges Papier Verfall nach 50-80 Jahren → Gefahr von Verlusten von nahezu 100% des aufgebauten Bestandes der letzten 150 Jahre
 - ▶ Maßnahme der Bestandserschließung
 - ▶ Lösung des Raumproblems

→ Gewinnung einer Kopie des gesamten Buches / Textes (spreche jetzt deshalb immer von Dokument), nach Urheberrecht mehr als nur ein Auszug, sondern ein wesentlicher Teil

kurz nochmal die wichtigsten Punkte des Urheberrechts

FOLIE

○ Urheberpersönlichkeitsrecht	<ul style="list-style-type: none">• Veröffentlichungsrecht• Anerkennung der Urheberschaft• Schutz gegen Entstellung des Werkes
○ Verwertungsrecht	<ul style="list-style-type: none">• Vervielfältigungsrecht• Verbreitungsrecht• Ausstellungsrecht • Recht der öffentlichen Wiedergabe• u.a.
○ Schutzdauer	<ul style="list-style-type: none">• erlischt 70 Jahre nach Tod des Urhebers• bei Anonyma u. Pseudonyma 70 Jahre nach Veröffentlichung

→ genauere Angaben dazu unter: <http://www.jere-mias.de/biwi/urheb1.html>

erst nach Ablauf Schutzfrist Werk gemeinfrei, d.h. kann von jedermann verwertet und verbreitet werden

Persönlichkeitsrechte müssen gewahrt bleiben (Zitierung)

→ Digitalisierung gemeinfreier Werke und ihre Nutzung unproblematisch

1) Problematisch:

Feststellung Anteil gemeinfreier Werke an den vom Zerfall bedrohten Beständen (Grossteil gemeinfrei)

sichere Feststellung i.d.R. nur nach umständlicher und zeitaufwendiger Recherche des Todesdatums des Urhebers → hoher Aufwand u. hohe Kosten (Verfahren nur für Einzelfälle möglich)

→ i.d.R. Digitalisierung u. Verwendung elektronischer Texte nur unter Beachtung des Urheberrechts möglich

Digitalisierung Akt der Vervielfältigung (*Unter Vervielfältigung versteht man die Übertragung eines Werkes auf einen anderen Träger. § 16 UrhG*)

2) Problematisch:

Digitalisierung auf Lizenzbasis? → zu hoher Aufwand Ermittlung Autor o. seiner Rechtsnachfolger auch Verlage u. Verwertungsgesellschaft WORT nicht erlaubt Lizenzierung Digitalisierung
→ höchstens Digitalisierung durch gesetzlich Lizenz

In der derzeitigen Rechtsprechung gibt es zwei unterschiedliche gesetzliche Legitimierungsansprüche. (genaue Ausführung des Rechtsstreites ginge zu weit)

Vollständigkeit halber nur mal kurz nennen

a) „Archivprivileg“

b) Vervielfältigung eines Vergriffenen Werkes zum **eigenen** Gebrauch

„Bücher oder Zeitschriften sind vergriffen, wenn sie vom Verlag seit 2 Jahren nicht mehr geliefert werden können.“ → gilt nicht, wenn Werk vom Verlag in anderer Form geliefert werden kann (z.B. CD-ROM, Books on Demand)

Herausfinden ob Buch/ Zeitung vergriffen ist:

Bücher anhand VLB (<http://www.buchhandel.de>), ähnliche ausländische Verzeichnisse, z.B. Books out of Print (<http://www.booksoutofprint.com/>)

für Periodica Nachfrage bei Verlagen

3) Problematisch:

fraglich ob (digitale) Kopien auch von Benutzer öffentl. Zugänglicher Bibliotheken genutzt werden können

→ Ausgehen von „historischem Willen des Gesetzgebers“

→ Kopien von Bibliotheken ja nicht zum Selbstzweck angefertigt, sondern zum Verleih

unstrittig Verleih von Vervielfältigungsstücken bei Papierkopien, Mikrofiche, digitalen Kopien auf transportablen Speichermedien, z.B. CD-ROM, Diskette

4) Problematisch:

öffentliche Aufführung bei elektronischen Kopien, da diese auf einem Server liegen

→ hier keine Ausleihe (Verbreitungsrecht des Urhebers nicht tangiert, aber eventuelle das Bestimmungsrecht über Öffentliche Aufführung)

→ Wiedergabe am Computerbildschirm für eine Person → nicht öffentlich und damit nicht urheberrechtlich relevant

5) Problematisch

auf das Werk können über Netz mehrere Leute zugleich zugreifen

→ dies im Moment nach geltender Rechtsprechung keine öffentliche Vorführung

„Schutzbedürfnis des Urhebers für den Bereich der Netznutzung“ zunehmend in Frage gestellt

→ hier zugunsten mit künftigen Recht eine Änderung zu erwarten, wenn UrhG an den Richtlinienvorschlag der EG angeglichen wird

Abschließend lässt sich sagen, dass das geltende Urheberrecht sowohl die Digitalisierung (der gefährdeten Bestände) als auch eine umfassende Nutzung der Digitalisierungsprodukte (für Bibliotheken) erlaubt.

Ausblick

Zunehmende technische Möglichkeiten machen eine weitergefasste Definition wohl in Zukunft notwendig. Schon heute wird häufig folgendes bereits zum Teil auch als Volltext verstanden.

Folie

Unter Volltext wird verstanden:

Ein Artikel- oder Buchtext [Dokument], der in seiner Gesamtheit elektronisch verfügbar ist.³

Hier wird nicht mehr auf das gespeicherte Format des Textes Wert gelegt.

Durch eine zunehmende Abkehr vom Buch-Prinzip halte ich folgende Definition für günstiger. Mein Versuch einer Definition lautet:

Folie

Volltext ist eine Publikation, die Information enthält, die nicht primär als Referenzinformation dient, deren Verlinkung zur Orientierung innerhalb des Dokuments dient und deren Verlinkung nach Außen nur der Referenzierbarkeit gilt.

Texte werden immer stärker und logischer strukturiert. Hier wird zunehmend bibliothekarische Know-How benötigt. Neben der Erfassung kann die Bibliothek auch Hilfestellungen zur Veröffentlichung geben. Die Aufgabenstellungen zwischen Urheber, Verlagen und Bibliotheken werden sich verändern.

³ Bibliotheksfachwortschatz (Version v. 23. Mai 2001), <http://www.lib.byu.edu/~lui/lingo/german.html>, Stand: 05.06.02