

Überprüfung der Spiegelungsfähigkeit von Internetressourcen mit dem Online-Tool ArchiveReady und Bewertung der dabei analysierten Facetten

Stephanie Kuch
Deutsches Literaturarchiv Marbach
kuch@dla-marbach.de

02. Mai 2016

Zusammenfassung

Im DFG-geförderten Projekt "Aufbau eines Quellencorpus für die seit den 1990er Jahren entstehende Literaturgattung „Netzliteratur“" wurde das Online-Tool „ArchiveReady“ von Vangelis Banos zur Vorabanalyse der zu archivierenden Werke verwendet. Das Tool überprüft die Archivtauglichkeit einer gegebenen Internetressource. Dieses Papier wird den Vorgang der Analyse im Projekt schildern, die Ergebnisse der Analyse darlegen und mit der erzielten Qualität der Spiegelungen in Zusammenhang setzen. Am Ende werden die überprüften Eigenschaften bezüglich ihre Auswirkung beim Archivierungsprozess im Projekt bewertet.

Das Projekt

Das DFG-geförderte Projekt "Aufbau eines Quellencorpus für die seit den 1990er Jahren entstehende Literaturgattung „Netzliteratur“" am Deutschen Literaturarchiv Marbach hatte es sich zwischen 2013 und 2016 zur Aufgabe gemacht, einen repräsentativen Korpus an Werken der Netzliteratur aus der Anfangszeit des World Wide Web zu archivieren.

Netzliteratur ist eine Literaturgattung, die nach der Freigabe der öffentlichen Benutzung des World Wide Web 1991 entstand. In der Netzliteratur fungiert das World Wide Web nicht nur als Publikationsform, sondern die (Hyperlink-)Struktur des WWW wird als Gestaltungselement und als Gegenstand ästhetischer Reflexion eingesetzt. Gerade diese Eigenschaft macht die Netzliteratur zu einem schwierig zu archivierenden Genre, da ein Medienwechsel (z. B. vom Internet auf das Papier) so gut wie unmöglich ist.

Der im Projekt ausgewählte Korpus umfasst 70 Werke, von denen für 50 die Erlaubnis zur Archivierung von den Rechteinhabern eingeräumt wurde. Von diesen 50 Werken waren 45 online verfügbar, sodass sie potentiell zur Archivierung mittels eines Webcrawlers geeignet sein würden. Unter den Werken befinden sich sowohl relativ einfache, rein HTML-basierte Websites als auch komplexe, die Javascript, Flash und anderes verwenden.

Die Archivierung von Netzliteratur kann nicht in allen Fällen mit den aktuell verbreiteten Webcrawlern adäquat durchgeführt werden. Daher müssen andere Methoden gefunden werden, um das Werk zu archivieren und/oder zu dokumentieren.

Im Projekt standen verschiedene Methoden zur Auswahl, die dem Zweck der Archivierung oder Dokumentation dienen.

- Methoden zur Dokumentation:
 - Screenshot
 - Screencast
- Methoden zur Archivierung:
 - Spiegelung (Crawl mittels eines Webcrawlers)
 - Archivierung des serverseitigen Codes und evtl. vorhandener Datenbanken

Von jedem zu archivierenden Werk wurde mindestens ein Screenshot angefertigt, um das aktuelle Erscheinungsbild zu dokumentieren. War keine ausreichende Spiegelung zu erreichen, so hatte die Dokumentation dieses Werks eine höhere Priorität.

Um den Archivierungsprozess effizienter zu gestalten und für jedes Werk die richtige Methode zu wählen wurden die Werke vorab auf ihre Spiegelungsfähigkeit analysiert.

Die Vorabanalyse

Für die Vorabanalyse wurde das Online-Tool ArchiveReady¹ verwendet. Das Tool basiert auf dem Papier „CLEAR – a credible method to evaluate website archivability“³ von Banos et al., das Facetten von Webseiten benennt, die den Erfolg einer Archivierung beeinflussen können. ArchiveReady überprüft Webseiten auf die im Papier benannten Eigenschaften.

Die dabei überprüften und gewichteten Facetten sind:

1. Zugänglichkeit (*Accessibility*)
2. Kohäsion (*Cohesion*)
3. Metadaten (*Metadata*)
4. Standardkonformität (*Standards Compliance*).

Diesen Facetten lassen sich die Reiter „HTML/CSS“, „HTTP“, „Media“ und „Sitemaps“ im Tool zuordnen.

¹ online verfügbar unter <http://archiveready.com/>, abgerufen am 26.01.2016. Die private Nutzung ist kostenlos. Bei größeren Mengen und Supportbedarf wird eine Gebühr fällig.

³ Banos, Vangelis. Kim, Yunhyong. Ross, Seamus. Manolopoulos, Yannis: CLEAR: a credible method to evaluate website archivability. 10th International Conference on Preservation of Digital Objects (iPres'2013), Lissabon, 2013.

ArchiveReady bewertet jede Facette einzeln und ermittelt aus den Einzelbewertungen eine Gesamteinschätzung der Archivtauglichkeit eines Werks, die einen Rückschluss auf die Spiegelungsfähigkeit des Werks zulässt. Neben einer Zusammenfassung des Ergebnisses kann über die Weboberfläche des Tools auch eine detaillierte Anzeige der Testergebnisse aufgerufen werden. Ein Beispiel dafür findet sich in Abbildung 1. Ein Link unterhalb des Testergebnisses (*Reference*) führt meist auf eine detaillierte Fehlerbeschreibung.

The screenshot displays the ArchiveReady website archiving evaluation tool interface. At the top, the tool's name and logo are visible, along with navigation links for Help, FAQ, and API. The main heading indicates the website being checked: `http://auer.netzliteratur.net/kill/killpoem.htm`. Below this, there is a search bar with the same URL and a 'Check again' button. A summary bar shows the following test results: HTML and CSS (4 errors), HTTP (3), Media (3), and Sitemaps (3). The main content area displays a list of test results:

- 01. Invalid HTML** `http://auer.netzliteratur.net/kill/killpoem.htm`. Located 16 errors, 4 warnings. Web pages which do not conform with Web Standards have a lower possibility to be preserved correctly. Rating: 0%, Significance: ★★★, Standards Compliance: -.
- 02. No CSS files used in this webpage.** The absence of CSS simplifies archiving. Rating: 100%, Significance: ★★★, Standards Compliance: +.
- 03. Local scripts found: 0, remote scripts found: 0** scripts used in a website but hosted elsewhere may cause problems in web archiving because they may not be captured when the site is archived. What is more, the site is depended on 3rd party sites which may not be available in the future, causing incontrolable problems. Rating: 100%, Significance: ★★, Cohesion: +.
- 04. No inline script elements found.** Inline JavaScript code inside HTML is considered a bad practice. The absence of inline JavaScript simplifies archiving. Rating: 100%, Significance: ★★★, Accessibility: +.

At the bottom, there are two sections: 'Linking to this result' which provides a URL for linking to the validation result, and 'Banner in your website' which provides HTML code to add a banner to the website.

Abbildung 1: Übersichtsseite der Testgruppe "HTML und CSS" in ArchiveReady, angefertigt am 26.01.2016

Mit diesen Angaben können fundierte Antworten auf folgende Fragen gefunden werden:

- Ist es wahrscheinlich, dass die Ressource zufriedenstellend mit einem Crawler archiviert werden kann?
- Welche Bereiche sind problematisch?

- Welche URLs müssen einem Crawler zusätzlich zur URL der Ressource übergeben werden, um alle relevanten Inhalte zu erhalten?
- Welche URLs sind nicht valide und können somit auch nicht archiviert werden?
- Wurde JavaScript gefunden?
- Wird Flash verwendet?
- Wie schnell antwortet der Server?
- Verhindert die robots.txt evtl. die Archivierung?
- Gibt es eine Sitemap?

ArchiveReady kann jedoch nicht überprüfen, ob der Inhalt einer Webseite dynamisch generiert wird oder ob Java Applications verwendet wurden. Dies wurde im Projekt manuell überprüft.

Um über die adäquate Archivierung eines Werks zu entscheiden, müssen diese Fragen selbstverständlich nicht nur für eine zum Werk gehörige Webseite beantwortet werden, sondern für alle Ressourcen, die zum Werk gehören. Da ArchiveReady jedoch immer nur die über die angegebene URL referenzierte Ressource überprüft und keinen Links folgt, müssen daher alle URLs, die zu einem Werk gehören an ArchiveReady übergeben werden.

Hierfür wurde im Projekt eine Linkliste erstellt, die alle URLs der Werke enthält. Die Linkliste wurde mittels `wget`² erstellt und durch manuelle Überprüfungen in einzelnen Fällen ergänzt, in denen eine automatisierte Linkextraktion nicht möglich war. Insgesamt wurden 3458 URLs identifiziert. Die gesammelten Links wurden über ein API an ArchiveReady übergeben. Das verwendete Verfahren wird detailliert im 2015 erschienenen Artikel „Praxisreport: Verfahren zur Evaluierung der Archivierbarkeit von Webobjekten“³ beschrieben. Am Ende des Prozesses stand pro Einzelwebseite eine Einzelbeurteilung. Diese Einzelbeurteilungen mussten für jedes Werk zusammengeführt werden, sodass ein vollständiger Bericht pro Werk entstand. Hierbei legte das schlechteste Testergebnis den Wert der Spiegelungsfähigkeit im Bericht fest. Ein Muster für einen zusammengeführten Bericht kann in Anhang 1 gefunden werden.

² online unter: <http://www.gnu.org/software/wget>

³ Fritz, Steffen: Praxisreport: Verfahren zur Evaluierung der Archivierbarkeit von Webobjekten. In: ABI Technik Zeitschrift für Automation, Bau und Technik im Archiv-, Bibliotheks- und Informationswesen, 2015, 35 (2), S. 117 - 120

Ergebnisse

Folgende Tabelle gibt eine quantitative Übersicht über die Werke, bei denen problematische Testergebnisse auftraten.

Eigenschaft	Anzahl der Werke
Invalides HTML/CSS	45
Kein (RSS-)Feed vorhanden	42
Keine Sitemap vorhanden	36
Verwendung von Javascript	28
Dynamisch generierte Inhalte*	20
Verwendung von Flash/Quicktime (Proprietäre Formate)	20
Verwendung von externen Inhalten (GoogleMaps, Google StreetView, aber auch Bilder, Videos, etc.)	11
Robots.txt, die Inhalte ausschließt	6
Invalide Hyperlinks	4
Invalide Medien	2
Verwendung von Java-Applets*	2

Tabelle 1: Quantitative Auswertung von problematischen Testergebnissen im Projekt

Die mit einem * markierten Eigenschaften sind nicht Teil der Analyse durch ArchiveReady, sondern wurden manuell im Zuge der Berichterstellung ermittelt.

Stichproben ergaben, dass ab einem Wert von 80% eine zufriedenstellende Spiegelung zu erwarten war. Bei Werken, die unter 80% fielen, wurde erwartet, dass bei einem Crawl entweder alle oder einige wichtige Eigenschaften des jeweiligen Werks verloren gehen würden, sodass hier die Anfertigung eines Screenscasts in jeden Fall nötig wäre, um die Funktions- und Erscheinungsweise des Werks zu dokumentieren.

Die Zusammenfassung aller Eigenschaften pro Werk ergab, dass

- 14 Werke mit 80% oder mehr ein sehr gutes Ergebnis erzielten,
- 28 Werke ein Ergebnis unter 80% erzielten,
- 3 Werke kein Ergebnis erzielten, da sie von ArchiveReady nicht korrekt analysiert wurden.

Nach Abschluss der Vorabanalyse wurden im Projekt Spiegelungen der Werke erstellt. Die Ergebnisse der Analyse mit ArchiveReady lieferten dabei Hinweise auf extern gehostete Inhalte, invalide Links und einzuschließende Domains. Die Spiegelungen wurden anschließend manuell auf ihre Qualität hin überprüft. Dies geschah an Hand einer im Projekt entwickelten Matrix, die in Tabelle 2 aufgeführt wird.

Basis-funktionalität	Name	Beschreibung	ja	nein	nicht getestet	nicht vorhanden
	Zugänglich	Kann die Spiegelung abgerufen werden?				
	Navigation	Funktioniert die Navigation korrekt?				
	Navigation eingeschränkt	Falls nein: Funktioniert die Navigation eingeschränkt?				
	Verlinkungen korrekt	Weisen alle Verlinkungen ausschließlich auf die archivierte Website?				
	Struktur	Wird die Seitenstruktur korrekt angezeigt?				
	Struktur nicht korrekt	Falls nein: Kann der Inhalt der Seite verwendet werden?				
Inhalte						
	Text	Wurden alle Texte erfasst?				
	Formatierung	Wurde die Formatierung der Texte korrekt erfasst?				
	Grafik	Wurden alle Grafikdateien erfasst?				
	Audio	Wurden alle Audiodateien erfasst?				
	Video	Wurden alle Videodateien erfasst?				
	Sonstige	Wurden sonstige Inhalte korrekt und vollständig erfasst?				
	Downloads	Funktionieren die angebotenen Downloads?				

Tabelle 2: Matrix zur Beurteilung der Qualität einer Spiegelung

Überprüft wurde, ob die Spiegelung die erwartete Basisfunktionalität erreicht hatte. Ein negativer Befund bewirkte, dass die Spiegelung als unzureichend eingestuft wurde und somit ein Screencast zur Dokumentation erstellt werden musste. Gegebenenfalls wurde die Spiegelung auch wiederholt. Erreichte die Spiegelung jedoch die Basisfunktionalität wurde

überprüft, ob alle Inhalte archiviert wurden. War dies nicht der Fall, wurde auch hier ein Screencast angefertigt oder bei Bedarf die Spiegelung mit neuen Parametern wiederholt.

Bei der Auswertung ergab sich, dass

- bei 24 Werke Basisfunktionalität und Inhalt vollständig erfasst wurden,
- bei 8 Werken die Basisfunktionalität erreicht wurde, aber der Inhalt nicht vollständig erfasst wurde,
- bei 11 Werken die Basisfunktionalität nicht erreicht wurde,
- bei 2 Werken keine Spiegelung angefertigt wurde.

Die Ergebnisse der Qualitätskontrolle wurden mit den Ergebnissen der Vorabanalyse in Relation gesetzt. Dabei stellt sich heraus, dass

- Werke, deren Spiegelungen sowohl die Basisfunktionalität als auch den Inhalt vollständig erfassten, in der Analyse mit ArchiveReady einen Gesamtwert zwischen 54% und 92% erreicht hatten. Der Mittelwert betrug 77,5 %.
- Werke, deren Spiegelungen nur die Basisfunktionalität erreichten, einen Gesamtwert zwischen 68 % und 87 % erreicht hatten. Der Mittelwert betrug 75,5 %.
- Werke, deren Spiegelungen als unzureichend eingestuft wurden, erreichten einen Gesamtwert zwischen 59 % und 78 %. Der Mittelwert betrug 72,1 %.

Die Ergebnisse zeigten, dass zwar eine Annahme über die Spiegelungsfähigkeit eines Werks aus dem von ArchiveReady bereitgestellten Gesamtwert abgeleitet werden kann, dies jedoch ein sehr unsicheres Verfahren ist. Geschuldet ist dies der Tatsache, das ArchiveReady auch Aspekte überprüft und gewichtet, die einen Einfluß auf die Langzeitarchivierbarkeit einer Ressource haben. Diese Aspekte gefährden die langfristige Verfügbarkeit einer Ressource, verhindern aber nicht eine Spiegelung und Bereitstellung mit heutigen Mitteln.

In einem weiteren Schritt wurden deshalb die der Ergebnisse von ArchiveReady und der anschließenden Qualitätskontrolle verglichen und überprüft, welche Komponenten für unzureichende oder fehlerhafte Spiegelungen verantwortlichen waren. Dies ermöglichte eine projektspezifische, differenzierte Bewertung der durch ArchiveReady überprüften Facetten.

Bewertung der durch ArchiveReady überprüften Facetten im Projekt

Im Folgenden soll geschildert werden, ob und welche Konsequenzen der negative Befund einer Eigenschaft auf die Spiegelungsfähigkeit des jeweiligen Werks im Projekt hatte.

- *Invalides HTML/CSS*: Diese Eigenschaft hatte keine Auswirkungen auf die Spiegelungsfähigkeit eines Werks. Crawler sind selbstverständlich auf korrekte Auszeichnung ausgerichtet und unter Umständen kann es vorkommen, dass Inhalte auf Grund von fehlerhafter Auszeichnung nicht erfasst werden. Praktisch kam dies im Projekt jedoch nicht vor. Bemerkenswert ist, dass keines der analysierten Werke über valide HTML- bzw. CSS-Auszeichnung verfügte.
- *Kein (RSS-)Feed vorhanden*: Diese Eigenschaft hatte keine Auswirkung auf die Spiegelungsfähigkeit eines Werks. Das Genre der Netzliteratur ist größtenteils abgeschlossen und die einzelnen Werke erfahren so gut wie keine Veränderungen mehr, sodass ein RSS-Feed bei der Ermittlung aktualisierter Inhalte nicht von Nutzen ist.
- *Verwendung von Flash/Quicktime*: Diese Eigenschaft hatte keine Auswirkung auf die Spiegelungsfähigkeit eines Werkes. Flash und Quicktime sind proprietäre Datenformate und deshalb nicht zur Langzeitarchivierung geeignet. Sie sind nicht quelloffen und benötigen spezielle Software um die Daten abspielen zu können. Somit ist die zukünftige Bereitstellung solcher Daten gefährdet. Doch der Spiegelungsvorgang an sich ist davon unberührt. Die archivierten Daten können heute noch zuverlässig wiedergegeben werden.
- *Invalide Hyperlinks*: Diese Eigenschaft hatte keine Auswirkung auf die Spiegelungsfähigkeit eines Werks. Invalide Hyperlinks führen auf eine Fehlerseite. Die Inhalte, die dort einmal hinterlegt waren existieren entweder nicht mehr oder wurden umgezogen. In jedem Fall kann dort keine Archivierung der ursprünglichen Inhalte stattfinden. Dies führt dazu, dass die archivierte Version nicht dem ursprünglichen Zustand entspricht und unvollständig ist. Abhilfe kann hier nur die Kommunikation mit dem Betreiber der Website schaffen. Eine Spiegelung des Werks in seinem unvollständigen Zustand wird aber nicht verhindert.
- *Invalide Medien*: Diese Eigenschaft hatte keine Auswirkung auf die Spiegelungsfähigkeit eines Werks. Ist ein Medium invalide bedeutet dies, dass es nicht der für das Dateiformat vorliegenden Spezifikation entspricht. Dies erschwert wiederum die Langzeitarchivierung, verhindert oder erschwert aber keine Spiegelung.

- *Verwendung von Java-Applets:* Diese Eigenschaft hatte keine Auswirkung auf die Spiegelungsfähigkeit eines Werks. Java-Applets können in der Regel gut gespiegelt werden, machen aber bei der Wiedergabe in der Wayback Machine oft Probleme.
- *Keine Sitemap vorhanden:* Diese Eigenschaft hatte eine geringe Auswirkung auf die Spiegelungsfähigkeit eines Werks. Generell hilft eine Sitemap dem Crawler zum Werk gehörige Links zu identifizieren. Dies wird besonders wichtig, wenn das Werk Skriptsprachen, wie beispielsweise Javascript enthält. Dort ist die Linkextraktion besonders schwierig. Eine Sitemap kann in solchen Fällen fehlende Links ergänzen.
- *Robots.txt, die Inhalte vom Spiegelungsvorgang ausschließt:* Diese Eigenschaft hatte eine geringe Auswirkung auf die Spiegelungsfähigkeit eines Werks. Die robots.txt teilt einem Crawler mit, ob und welcher Inhalte er erfassen darf. Problematisch wird es, wenn Inhalte, die essentiell für das Werk sind, ausgeschlossen sind oder gar der für die Spiegelung verwendete Crawler selbst ausgeschlossen ist.
- *Verwendung von (inline) Javascript:* Diese Eigenschaft hatte eine große Auswirkung auf die Spiegelungsfähigkeit eines Werks. Generell sind Crawler auf die Bearbeitung von HTML-Seiten ausgelegt. In Javascript wird aktuell nur nach Strings gescannt, die den Mustern von absoluten oder relativen Links folgen. Diese werden dann als vermutete URLs wieder in den Spiegelungsprozess aufgenommen⁴. Dieses Verfahren ist selbstverständlich fehleranfällig. Links werden deshalb oft nicht entdeckt.
- *Dynamisch generierte Inhalte:* Diese Eigenschaft hatte eine große Auswirkung auf die Spiegelungsfähigkeit eines Werks. Dynamisch generierte Websites enthalten Inhalte, die serverseitig erzeugt und dem Browser dann als HTML-Seite präsentiert werden. Die Website kann so bei jedem Aufruf anders aussehen. Das Problem hierbei ist, dass Inhalte, die nicht angezeigt werden auch nicht archiviert werden können. Bei einem Crawlvorgang wird somit nur eine von unendlich vielen Möglichkeiten archiviert.
- *Verwendung von externen Inhalten:* Diese Eigenschaft hatte eine große Auswirkung auf die Spiegelungsfähigkeit eines Werks. Ein Crawler bleibt zunächst immer innerhalb der ihm vorgegebenen Domain. Dies bedeutet, dass externe Inhalte vom Crawlvorgang standardmäßig ausgeschlossen und somit nicht gespiegelt werden. Handelt es sich um Bilder, Videos. etc. können mit den durch ArchiveReady ermittelten Angaben die externen Ressourcen in den Crawlvorgang eingeschlossen

⁴ siehe dazu: o.A. (2015): Crawling Javascript, in: Heritrix Wiki, online unter: <https://webarchive.jira.com/wiki/display/Heritrix/crawling+JavaScript>, abgerufen am 01.03.2016

werden. Handelt es sich jedoch um interaktive externe Inhalte, wie beispielsweise Google Maps oder StreetView, können diese Inhalte nicht archiviert werden.

Fazit

Die Verwendung von ArchiveReady als Voranalyse der zu archivierenden Werke war im Projekt sehr sinnvoll. So konnte durch die ermittelten Angaben der Spiegelungsvorgang effizienter gestaltet werden, da mehrfache Crawlvorgänge in den meisten Fällen vermieden werden konnten. Angaben zu externen Bestandteilen, robots.txt und invaliden Links waren dabei von unschätzbarem Wert.

Wie jedoch schon angedeutet, analysiert und gewichtet ArchiveReady sowohl Aspekte, die eine Ressource archivierungsfähig machen, als auch Aspekte, die eine Ressource archivtauglich machen. Zwischen diesen muss bei der Analyse und Interpretation der Ergebnisse deutlich unterschieden werden.

Die Verwendung von validen und quelloffenen Datenformaten ist sehr kritisch für eine gelungene Langzeitarchivierung, verhindert eine Spiegelung jedoch nicht. Ein Archiv muss in diesen Fällen abwägen, ob ein Werk dann trotzdem aufgenommen wird und wie mit den Bestandteilen, welche die Langzeitarchivierung gefährden, umzugehen ist.

Eine zukünftige Bereitstellung ist in solchen Fällen evtl. schwierig, da die Datenformate möglicherweise nicht mehr gelesen werden können bzw. die enthaltenen Informationen nicht mehr darstellbar sind, da die benötigte Hard- und Software nicht mehr verfügbar ist.

Im Projekt wurde dieser Problematik mit der Erstellung von Metadaten begegnet, die eine zukünftige Lesbarkeit der Daten wahrscheinlicher machen sollen. So wurde die zum Auslesen der Daten verwendete Hard- und Software dokumentiert, sodass in Zukunft die Umgebung wieder nachgebaut werden könnte.

Somit lassen sich mit ArchiveReady zwei wichtige Zweige der Archivierung von Internetressourcen abdecken: der Archivierungsvorgang an sich mit seinen heutigen Herausforderungen und die langfristige Bereitstellung und Archivierung dieser Ressourcen. Um das Tool jedoch erfolgreich nutzen zu können, sollte man sich die Bedeutung und Konsequenzen der einzelnen Eigenschaften klar vor Augen führen und neben dem Gesamtwert immer die Ergebnisse der Einteltests analysieren. Im Idealfall kann mit Hilfe von ArchiveReady eine vollständige Archivierung einer Netzressource gelingen, die Sicherstellung einer authentischen Darstellung der archivierten Daten obliegt jedoch weiterhin der archivierenden Institution. Dazu müssen benötigte Hard- und Software beschrieben und

die als authentisch erachtete Darstellung dokumentiert werden. Dies sollte mittels einer ausführlichen Beschreibung der Hard- und Softwareumgebung in den Metadaten zum Objekt geschehen. Zusätzlich ist eine Dokumentation des authentischen Erscheinungsbilds des Werks mittels Screenshots und Screencasts zu empfehlen.

Anhang 1

Beurteilung Spiegelungsfähigkeit

Beschreibung Abspielumgebung

Werktitel	
Werkversion	
Autor	
Hauptseite-URL	
Datum der Sichtung	
Bearbeiter	
Einrichtung	
Betriebssystem	
Browser	
Plugins	

Eine Liste aller zum Werk gehörenden und getesteten URLs findet sich im Anhang A. Das Ergebnis der automatisierten Evaluation mittels ArchiveReady findet sich in Anhang B.

Für die Beurteilung der Archivierbarkeit ist die Anzahl der negativ gewerteten Eigenschaften unerheblich.

Teilbereiche

Dieser Beurteilungsleitfaden beruht auf dem Paper „CLEAR: a credible method to evaluate website archivability“⁵ und dient als Grundlage für die Entwicklung eines Programms, das die Beurteilung semiautomatisch durchführt.

Zugänglichkeit

1. Validieren alle Hyperlinks? - Ja/Nein.
2. Werden Inhalte dynamisch generiert? - Ja/Nein.
3. Gibt es bereits eine archivierte Version, z. B. bei archive.org? Welche Qualität hat diese? - Ja/Nein.
4. Verfügt die Seite über eine Sitemap? - Ja/Nein.
5. Verfügt die Seite über eine robots.txt? Schließt diese Crawler oder bestimmten Content aus? - Ja/Nein.
6. Verfügt die Seite über Feeds? - Ja/Nein.
7. Werden proprietäre Datenformate (Flash/Quicktime) verwendet? - Ja/Nein.

Standardkonformität

1. Validierung referenzierter Medien (Bilder, Audio, Video, ...) - (Nicht) bestanden.

⁵ Banos et al.: *CLEAR: a credible method to evaluate website archivability*. Lisbon. 2013.

2. Compliance-Prüfung von Webelementen (HTML, CSS) - (Nicht) bestanden.
3. Compliance-Prüfung ergänzender Ressourcen (sitemap.xml, robots.txt, ...) - (Nicht) bestanden.

Performanz

Werden Anfragen ausreichend schnell beantwortet? - Ja/Nein.

Kohäsion

1. Werden Dateien auf verschiedenen Systemen gehostet? - Ja/Nein.

Metadaten

Sind Metadaten aus HTTP Headern vorhanden? - Ja/Nein.

Bemerkungen

Beurteilung

Der schlechteste Wert aus der Evaluierung aller Dokumente ist maßgeblich für die Archivierbarkeit der gesamten Webressource.

Archivierbarkeit xx%

Anhang A: Liste aller zum Werk gehörenden und getesteten URLs

Anhang B