

## **Beurteilung Spiegelungsfähigkeit**

### **Beschreibung Abspielumgebung**

|                    |   |
|--------------------|---|
| Werktitel          | Streifen-Generator  |
| Werkversion        | 20.11.2014  |
| Autor              | Manfred Arens   |
| Hauptseite-URL     | <a href="http://www.untex.de/generatoren/streifen.htm">http://www.untex.de/generatoren/streifen.htm</a> |
| Datum der Sichtung | 20.11.2014  |
| Bearbeiter         | S. Kuch   |
| Einrichtung        | Deutsches Literaturarchiv Marbach   |
| Betriebssystem     | Windows XP  |
| Browser            | Firefox 17.0.7  |
| Plugins            | Java Plugin 10.65.2.20  |

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“<sup>1</sup> und dient als Grundlage für die Entwicklung eines Programms, das die Beurteilung semiautomatisch durchführt.

### **Zugänglichkeit**

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

### **Standardkonformität**

1. Validierung referenzierter Medien (Bilder, Audio, Video, ...) - Keine vorhanden.
2. Compliance-Prüfung von Webelementen (HTML, CSS) - nicht bestanden (invalides HTML)

### **Performanz**

Werden Anfragen ausreichend schnell beantwortet? - Nicht bestanden.

### **Kohäsion**

Werden Dateien auf verschiedenen Systemen gehostet? - Nein.

---

<sup>1</sup> Banos et al.: *CLEAR: a credible method to evaluate website archivability*. Lisabon. 2013.

## Metadaten

Sind Metadaten aus HTTP Headern vorhanden? - Nein.

## Bemerkungen

Die Archivierbarkeit bezieht sich auf statische Inhalte, die allerdings dynamisch generiert werden.

## Beurteilung

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

Archivierbarkeit 77%

### Anhang A: Liste aller zum Werk gehörenden und getesteten URLs

<http://www.untex.de/generatoren/streifen1.htm>  
[http://www.untex.de/generatoren/streifen\\_c.htm](http://www.untex.de/generatoren/streifen_c.htm)  
<http://www.untex.de/generatoren/streifen.htm>  
[http://www.untex.de/generatoren/streifen\\_i.htm](http://www.untex.de/generatoren/streifen_i.htm)

### Anhang B

URL: <http://www.untex.de/generatoren/streifen.htm>  
Testdatum: 2014-11-11 12:40:10.979752  
Standard Compliance: 69  
Archivierbarkeit: 81  
Accessibility: 56  
Kohäsion: 100

+++ Meldungen +++

-----  
Titel: No sitemap.xml found.  
Level: 0  
Gewichtung: HIGH  
Beschreibung:  
Sitemaps are an easy way for webmasters to inform web bots about available web pages so that they can more intelligently crawl the site.

-----  
Titel: No images found in this page.  
Level: 100  
Gewichtung: MEDIUM  
Beschreibung:  
Therefore, it is easier to be archived.

-----  
Titel: You are not using any external proprietary objects (Flash, Quicktime)  
Level: 100  
Gewichtung: HIGH  
Beschreibung:  
Proprietary formats are not recommended for , long term preservation.

Please consider changing , your content to open formats.

-----  
Titel: Invalid HTML <http://www.untex.de/generatoren/streifen.htm>. Located 19 errors, 2 warnings.

Level: 0

Gewichtung: HIGH

Beschreibung:

Web pages which do not conform with Web Standards have a lower possibility to be preserved correctly

-----  
Titel: No CSS files used in this webpage.

Level: 100

Gewichtung: HIGH

Beschreibung:

The absence of CSS simplifies archiving.

-----  
Titel: Local scripts found: 0, remote scripts found: 0

Level: 100

Gewichtung: MEDIUM

Beschreibung:

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 incontrollable problems.<br />

-----  
Titel: No inline script elements found

Level: 100

Gewichtung: HIGH

Beschreibung:

Inline JavaScript code inside HTML is considered a bad practice. The absence of inline JavaScript simplifies archiving.

-----  
Titel: No Content-Encoding or Transfer-Encoding defined in HTTP Headers.

Level: 50

Gewichtung: MEDIUM

Beschreibung:

The encoding used on the data is not defined. Web bots , would try assume the possible encoding, leading to potential , mistakes. A Transfer-Encoding header field, which can be used , to specify an auxiliary encoding that was applied , to the data in order to allow it to pass through mail , transport mechanisms , which may have data or character set limitations.

-----  
Titel: Content type was clearly defined.

Level: 100

Gewichtung: MEDIUM

Beschreibung:

Correctly defined HTTP headers help web bots understand the content.

-----  
Titel: HTTP caching headers enabled.

Level: 100

Gewichtung: MEDIUM

Beschreibung:

HTTP Expires: was not defined.<br />HTTP Last-Modified: was clearly defined

in HTTP headers: Thu, 02 Aug 2007 14:50:23 GMT<br />HTTP Etag was clearly defined in HTTP headers: "e0513d4c-2cb-436b88fe6e9c0"<br />HTTP caching headers are critical to improve web archiving bots performance because they help avoid downloading unmodified content multiple times.

-----  
Titel: Network response time is 3000 ms  
Level: 0  
Gewichtung: HIGH  
Beschreibung:  
Website performance affects greatly the process of data aggregation and web accessibility. Performance evaluation rating is 100 when response time is 200ms or less and 0 when response time is more than 2sec.

URL: [http://www.untexte.de/generatoren/streifen\\_c.htm](http://www.untexte.de/generatoren/streifen_c.htm)  
Testdatum: 2014-11-11 12:40:09.153008  
Standard Compliance: 69  
Archivierbarkeit: 81  
Accessibility: 56  
Kohäsion: 100

+++ Meldungen +++

-----  
Titel: No sitemap.xml found.  
Level: 0  
Gewichtung: HIGH  
Beschreibung:  
Sitemaps are an easy way for webmasters to inform web bots about available web pages so that they can more intelligently crawl the site.

-----  
Titel: No images found in this page.  
Level: 100  
Gewichtung: MEDIUM  
Beschreibung:  
Therefore, it is easier to be archived.

-----  
Titel: You are not using any external proprietary objects (Flash, Quicktime)  
Level: 100  
Gewichtung: HIGH  
Beschreibung:  
Proprietary formats are not recommended for , long term preservation. Please consider changing , your content to open formats.

-----  
Titel: Invalid HTML [http://www.untexte.de/generatoren/streifen\\_c.htm](http://www.untexte.de/generatoren/streifen_c.htm).  
Located 4 errors, 2 warnings.  
Level: 0  
Gewichtung: HIGH  
Beschreibung:  
Web pages which do not conform with Web Standards have a lower possibility to be preserved correctly

-----  
Titel: No CSS files used in this webpage.  
Level: 100  
Gewichtung: HIGH

Beschreibung:  
The absence of CSS simplifies archiving.

-----  
Titel: Local scripts found: 0, remote scripts found: 0  
Level: 100  
Gewichtung: MEDIUM

Beschreibung:  
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 incontrollable problems.<br />

-----  
Titel: No inline script elements found  
Level: 100  
Gewichtung: HIGH

Beschreibung:  
Inline JavaScript code inside HTML is considered a bad practice. The absence of inline JavaScript simplifies archiving.

-----  
Titel: No Content-Encoding or Transfer-Encoding defined in HTTP Headers.  
Level: 50  
Gewichtung: MEDIUM

Beschreibung:  
The encoding used on the data is not defined. Web bots , would try assume the possible encoding, leading to potential , mistakes. A Transfer-Encoding header field, which can be used , to specify an auxiliary encoding that was applied , to the data in order to allow it to pass through mail , transport mechanisms , which may have data or character set limitations.

-----  
Titel: Content type was clearly defined.  
Level: 100  
Gewichtung: MEDIUM

Beschreibung:  
Correctly defined HTTP headers help web bots understand the content.

-----  
Titel: HTTP caching headers enabled.  
Level: 100  
Gewichtung: MEDIUM

Beschreibung:  
HTTP Expires: was not defined.<br />HTTP Last-Modified: was clearly defined in HTTP headers: Thu, 02 Aug 2007 14:50:24 GMT<br />HTTP Etag was clearly defined in HTTP headers: "e0513d5b-562-436b88ff62c00"<br />HTTP caching headers are critical to improve web archiving bots performance because they help avoid downloading unmodified content multiple times.

-----  
Titel: Network response time is 3000 ms  
Level: 0  
Gewichtung: HIGH

Beschreibung:  
Website performance affects greatly the process of data aggregation and web accessibility. Performance evaluation rating is 100 when response time is 200ms or less and 0 when response time is more than 2sec.

URL: [http://www.untexte.de/generatoren/streifen\\_i.htm](http://www.untexte.de/generatoren/streifen_i.htm)  
Testdatum: 2014-11-11 12:40:12.524207  
Standard Compliance: 69  
Archivierbarkeit: 81  
Accessibility: 56  
Kohäsion: 100

+++ Meldungen +++

-----  
Titel: No sitemap.xml found.  
Level: 0  
Gewichtung: HIGH  
Beschreibung:  
Sitemaps are an easy way for webmasters to inform web bots about available web pages so that they can more intelligently crawl the site.

-----  
Titel: No images found in this page.  
Level: 100  
Gewichtung: MEDIUM  
Beschreibung:  
Therefore, it is easier to be archived.

-----  
Titel: You are not using any external proprietary objects (Flash, Quicktime)  
Level: 100  
Gewichtung: HIGH  
Beschreibung:  
Proprietary formats are not recommended for , long term preservation. Please consider changing , your content to open formats.

-----  
Titel: Invalid HTML [http://www.untexte.de/generatoren/streifen\\_i.htm](http://www.untexte.de/generatoren/streifen_i.htm).  
Located 3 errors, 2 warnings.  
Level: 0  
Gewichtung: HIGH  
Beschreibung:  
Web pages which do not conform with Web Standards have a lower possibility to be preserved correctly

-----  
Titel: No CSS files used in this webpage.  
Level: 100  
Gewichtung: HIGH  
Beschreibung:  
The absence of CSS simplifies archiving.

-----  
Titel: Local scripts found: 0, remote scripts found: 0  
Level: 100  
Gewichtung: MEDIUM  
Beschreibung:  
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 incontrollable problems.<br />

-----  
Titel: No inline script elements found

Level: 100

Gewichtung: HIGH

Beschreibung:

Inline JavaScript code inside HTML is considered a bad practice. The absence of inline JavaScript simplifies archiving.

-----  
Titel: No Content-Encoding or Transfer-Encoding defined in HTTP Headers.

Level: 50

Gewichtung: MEDIUM

Beschreibung:

The encoding used on the data is not defined. Web bots , would try assume the possible encoding, leading to potential , mistakes. A Transfer-Encoding header field, which can be used , to specify an auxiliary encoding that was applied , to the data in order to allow it to pass through mail , transport mechanisms , which may have data or character set limitations.

-----  
Titel: Content type was clearly defined.

Level: 100

Gewichtung: MEDIUM

Beschreibung:

Correctly defined HTTP headers help web bots understand the content.

-----  
Titel: HTTP caching headers enabled.

Level: 100

Gewichtung: MEDIUM

Beschreibung:

HTTP Expires: was not defined.<br />HTTP Last-Modified: was clearly defined in HTTP headers: Thu, 02 Aug 2007 14:50:24 GMT<br />HTTP Etag was clearly defined in HTTP headers: "e0513d5c-330-436b88ff62c00"<br />HTTP caching headers are critical to improve web archiving bots performance because they help avoid downloading unmodified content multiple times.

-----  
Titel: Network response time is 3000 ms

Level: 0

Gewichtung: HIGH

Beschreibung:

Website performance affects greatly the process of data aggregation and web accessibility. Performance evaluation rating is 100 when response time is 200ms or less and 0 when response time is more than 2sec.

URL: <http://www.untexte.de/generatoren/streifen1.htm>

Testdatum: 2014-11-11 12:40:07.481788

Standard Compliance: 69

Archivierbarkeit: 77

Accessibility: 39

Kohäsion: 100

+++ Meldungen +++

-----  
Titel: No sitemap.xml found.

Level: 0

Gewichtung: HIGH

Beschreibung:

Sitemaps are an easy way for webmasters to inform web bots about available web pages so that they can more intelligently crawl the site.

-----  
Titel: No images found in this page.  
Level: 100  
Gewichtung: MEDIUM  
Beschreibung:  
Therefore, it is easier to be archived.

-----  
Titel: You are not using any external proprietary objects (Flash, Quicktime)  
Level: 100  
Gewichtung: HIGH  
Beschreibung:  
Proprietary formats are not recommended for , long term preservation.  
Please consider changing , your content to open formats.

-----  
Titel: Invalid HTML <http://www.untex.de/generatoren/streifen1.htm>.  
Located 29 errors, 2 warnings.  
Level: 0  
Gewichtung: HIGH  
Beschreibung:  
Web pages which do not conform with Web Standards have a lower possibility to be preserved correctly

-----  
Titel: There are 6 valid and 3 invalid links.  
Level: 66  
Gewichtung: HIGH  
Beschreibung:  
Invalid links prevent web bots from accessing the full content , in order to preserve it. Also check out [http://en.wikipedia.org/wiki/Link\\_rot](http://en.wikipedia.org/wiki/Link_rot) , <br />Invalid links:<br /> , <http://www.untex.de/generatoren/streifen1.htm>/  
<http://www.untex.de/generatoren/streifen1.htm>:100<br /><http://www.untex.de/generatoren/streifen1.htm/>  
<http://www.untex.de/generatoren/streifen1.htm>:180 in codieren<br />TypeError: 'undefined' is not an object (evaluating 'parent.frame2.document')<br /> , Valid links:<br /> , <http://www.untex.de/index.htm><br /><http://www.untex.de/generatoren.htm><br /><http://www.untex.de/untex/interior.htm><br />[http://www.untex.de/generatoren/streifen\\_c.htm](http://www.untex.de/generatoren/streifen_c.htm)<br />[http://www.untex.de/generatoren/streifen\\_i.htm](http://www.untex.de/generatoren/streifen_i.htm)<br /><http://www.untex.de/generatoren/dreh.htm><br />

-----  
Titel: No CSS files used in this webpage.  
Level: 100  
Gewichtung: HIGH  
Beschreibung:  
The absence of CSS simplifies archiving.

-----  
Titel: Local scripts found: 0, remote scripts found: 0  
Level: 100  
Gewichtung: MEDIUM  
Beschreibung:  
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 incontrollable problems.<br />

-----  
Titel: 3 inline script elements found  
Level: 0  
Gewichtung: HIGH  
Beschreibung:  
Inline Javascript code inside HTML is considered a bad practice.

-----  
Titel: No Content-Encoding or Transfer-Encoding defined in HTTP Headers.  
Level: 50  
Gewichtung: MEDIUM  
Beschreibung:  
The encoding used on the data is not defined. Web bots , would try assume the possible encoding, leading to potential , mistakes. A Transfer-Encoding header field, which can be used , to specify an auxiliary encoding that was applied , to the data in order to allow it to pass through mail , transport mechanisms , which may have data or character set limitations.

-----  
Titel: Content type was clearly defined.  
Level: 100  
Gewichtung: MEDIUM  
Beschreibung:  
Correctly defined HTTP headers help web bots understand the content.

-----  
Titel: HTTP caching headers enabled.  
Level: 100  
Gewichtung: MEDIUM  
Beschreibung:  
HTTP Expires: was not defined.<br />HTTP Last-Modified: was clearly defined in HTTP headers: Sun, 02 Nov 2014 17:37:14 GMT<br />HTTP Etag was clearly defined in HTTP headers: "e0513d58-247d-506e3ae4914a5"<br />HTTP caching headers are critical to improve web archiving bots performance because they help avoid downloading unmodified content multiple times.

-----  
Titel: Network response time is 3000 ms  
Level: 0  
Gewichtung: HIGH  
Beschreibung:  
Website performance affects greatly the process of data aggregation and web accessibility. Performance evaluation rating is 100 when response time is 200ms or less and 0 when response time is more than 2sec.