

Beurteilung Spiegelungsfähigkeit

Beschreibung Abspelunggebung

Werktitel	Fernwärme
Werkversion	18.11.2014
Autor	Ursula Hentschläger, Zelko Wiener
Hauptseite-URL	http://www.zeitgenossen.com/fernwaerme/
Datum der Sichtung	18.11.2014
Bearbeiter	S. Kuch
Einrichtung	Deutsches Literaturarchiv Marbach
Betriebssystem	Windows XP
Browser	Firefox 17.0.7
Plugins	Shockwave Flash 11.8.800.94

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

Zugänglichkeit

1. Validieren alle Hyperlinks? - Keine vorhanden.
2. Werden Inhalte dynamisch generiert? - Nein.
3. Gibt es bereits eine archivierte Version, z. B. bei archive.org? Welche Qualität hat diese? - Ja, funktionsfähig
4. Verfügt die Seite über eine Sitemap? - Nein.
5. Verfügt die Seite über eine robots.txt? Schließt diese Crawler oder bestimmten Content aus? - Ja. Ja.
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.

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

Metadaten

Sind Metadaten aus HTTP Headern vorhanden? - Nein.

Bemerkungen

Beurteilung

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

Archivierbarkeit 68%

Anhang A: Liste aller zum Werk gehörenden und getesteten URLs

<http://www.zeitgenossen.com/fernwaerme/>
<http://www.zeitgenossen.com/fernwaerme/indexin.html>
<http://www.zeitgenossen.com/fernwaerme/indexout.html>
<http://www.zeitgenossen.com/robots.txt>

Anhang B

URL: <http://www.zeitgenossen.com/fernwaerme/>
Testdatum: 2014-11-11 08:44:02.262996
Standard Compliance: 69
Archivierbarkeit: 79
Accessibility: 45
Kohäsion: 100

+++ Meldungen +++

Titel: There are "Disallow:" commands in your robots.txt file
Level: 0
Gewichtung: MEDIUM
Beschreibung:
Disallow:
Disallow: commands are used to block web bots from specific website areas. They may cause web archiving problems.

Titel: There is no sitemap.xml reference in your robots.txt file.
Level: 0
Gewichtung: MEDIUM
Beschreibung:
This can be used to tell search engines or other robots where your sitemap is located.

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.zeitgenossen.com/fernwaerme/>. Located errors, 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.

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.
HTTP Last-Modified: was clearly defined in HTTP headers: Mon, 18 Sep 2006 10:08:15 GMT
HTTP Etag was clearly defined in HTTP headers: "804e39ce-491-41db78d60e1c0"
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.zeitgenossen.com/fernwaerme/indexin.html>
Testdatum: 2014-11-11 08:44:04.464253
Standard Compliance: 44
Archivierbarkeit: 68
Accessibility: 27
Kohäsion: 100

+++ Meldungen +++

Titel: There are "Disallow:" commands in your robots.txt file
Level: 0
Gewichtung: MEDIUM
Beschreibung:
Disallow:
Disallow: commands are used to block web bots from specific website areas. They may cause web archiving problems.

Titel: There is no sitemap.xml reference in your robots.txt file.
Level: 0
Gewichtung: MEDIUM
Beschreibung:
This can be used to tell search engines or other robots where your sitemap is located.

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 using proprietary external objects (Flash, Quicktime).
Level: 0
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.zeitgenossen.com/fernwaerme/indexin.html>.
Located errors, 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.

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.
HTTP Last-Modified: was clearly defined in HTTP headers: Mon, 18 Sep 2006 10:08:16 GMT
HTTP Etag was clearly defined in HTTP headers: "804e39cf-636-41db78d702400"
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.zeitgenossen.com/fernwaerme/indexout.html>
Testdatum: 2014-11-11 08:44:05.521016
Standard Compliance: 69
Archivierbarkeit: 79
Accessibility: 45
Kohäsion: 100

+++ Meldungen +++

Titel: There are "Disallow:" commands in your robots.txt file
Level: 0
Gewichtung: MEDIUM
Beschreibung:
Disallow:
Disallow: commands are used to block web bots from specific website areas. They may cause web archiving problems.

Titel: There is no sitemap.xml reference in your robots.txt file.
Level: 0
Gewichtung: MEDIUM
Beschreibung:
This can be used to tell search engines or other robots where your sitemap is located.

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.zeitgenossen.com/fernwaerme/indexout.html>.
Located errors, 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.

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.
HTTP Last-Modified: was clearly defined in HTTP headers: Mon, 18 Sep 2006 10:08:16 GMT
HTTP Etag was clearly defined in HTTP headers: "804e39d1-a4-41db78d702400"
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.zeitgenossen.com/robots.txt>
Testdatum: 2014-11-11 08:44:03.354696
Standard Compliance: 69
Archivierbarkeit: 79
Accessibility: 45
Kohäsion: 100

+++ Meldungen +++

Titel: There are "Disallow:" commands in your robots.txt file
Level: 0
Gewichtung: MEDIUM
Beschreibung:
Disallow:
Disallow: commands are used to block web bots from specific website areas. They may cause web archiving problems.

Titel: There is no sitemap.xml reference in your robots.txt file.
Level: 0
Gewichtung: MEDIUM
Beschreibung:
This can be used to tell search engines or other robots where your sitemap is located.

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.zeitgenossen.com/robots.txt>. Located errors, 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.

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.
HTTP Last-Modified: was clearly defined in HTTP headers: Tue, 21 Feb 2006 17:55:41 GMT
HTTP Etag was clearly defined in HTTP headers: "4072627b-17-40d51b8378140"
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.