

Webarchivierung in den National Archives UK

Vorteile von MirrorWeb



Jens Peters
Walter Nagel GmbH & Co. KG

Themen

- Vorstellung und Einführung

Zwischenfrage

- Warum sollten Sie sich MirrorWeb ansehen?
- Beispiel: The National Archives UK
- Und nun?

Vorstellung und Einführung

Walter Nagel GmbH & Co. KG

- Seit mehr als 40 Jahren arbeiten wir mit Bibliotheken, Archive und Museen
- Größter unabhängiger Partner für Kulturgutscanner (Deutschland, Österreich, Schweiz)
- Anbieter der meist genutzten Softwarelösung Visual Library für digitale Geschäftsgänge (z.B. Retrodigitalisierung und ePflicht)
- Webseiten- und Social-Media-Archivierung
- Langzeitarchivierung

Wir haben nichts zu verbergen!

www.walternagel.de/referenzen

- Bundesarchiv, Geheimes Staatsarchiv, Landes- und Staatsarchive NRW, Rheinland-Pfalz, Bayern, Thüringen, Schleswig-Holstein, Niedersachsen, Basel, Thurgau etc.
- Nahezu 100 Stadt- und Kreisarchive
- Staats- und Landesbibliotheken Berlin, München, Düsseldorf, Stuttgart, Darmstadt, Koblenz, Aargau, St. Gallen, Wien
- Universitätsbibliotheken Berlin, Augsburg, Köln, Basel, Leipzig, Linz, Klagenfurt, Frankfurt, Münster, Mainz
- Stadtmuseen Andernach, Spandau, Wewelsburg, Kevelaer
- Kunstmuseen Albertina Wien, Basel, Galerie Belvedere, MuMok

Vorstellung und Einführung

Das Unternehmen MirrorWeb

- Seit 2012 aktiv - mit Sitz in Manchester
 - Konzentration nur auf Webseitenarchivierung
 - Fokus auf Standards (ISO/WARC + SHA1)
 - Spezialisiert auf Entwicklung von Big-Data-Lösungen
 - Zu den Kunden gehören Banken, Versicherungen, Archive, Bibliotheken und Unternehmen aus der ganzen Welt
1. Treffen auf dem Archivtag in Glasgow
 2. Erst nach dem Feedback erfahrener „Webarchivierer“ entschlossen die Lösung anzubieten.

Vorstellung und Einführung

Die Lösung MirrorWeb

- Integrierte Plattform inkl. Volltextsuche über alle Inhalte, Statistiken, Automatisierung der Crawls
- Inkl. Hosting in einem modernen Datacenter in Deutschland
- Erweiterung des Heritrix-Crawlers + Eigenentwicklung „Electrolyte“
- ISO-konform (WARC)

Vorstellung und Einführung



BANK OF ENGLAND



unicef



UNHCR
The UN Refugee Agency



Llywodraeth Cymru
Welsh Government



Department
for Exiting the
European Union



HOUSE OF LORDS



HOUSES OF PARLIAMENT



UNIVERSITY OF
WESTMINSTER



Arro



Bewahren Sie einzelne Seiten oder ganze Bücher?

INTERNET ARCHIVE
waybackmachine

http://www.walternagel.de/index.php Go JUL AUG NOV
2003 26 2004 2005

16 captures
28 Aug 2004 - 3 Apr 2015

Walter Nagel

Home
Sitemap
Kontakt
Impressum

Produkte Unternehmen News Partner

Herzlich willkommen

Organisation im Büro- und Gesundheitswesen

Wir freuen uns, dass Sie uns über diese Seiten kennen lernen möchten. Hier erfahren Sie alles Wissenswerte über unser Unternehmen, das Produktangebot und unsere Partner. Wir wünschen Ihnen viel Vergnügen.

Schon gewusst?
Kennen Sie das Handbuch zur HINZ-Patientendokumentation für den Heim Bereich

Wenn Sie aktuelle Produktinformationen wünschen, melden Sie sich bitte hier mit Ihrer E-Mail Adresse an:

Nancy
NANCY ermöglicht die Erstellung von umfassenden Standardpflegeplänen und Patientendokumentation unter Beachtung von Diagnosen aus Pflege und Medizin. [mehr..](#)

BeAMflow
Mit dem Designer-Modul steht jetzt ein grafisches Tool zur Konfiguration der kompletten Dokumentationssysteme zur Verfügung. [mehr..](#)

Starke Partner
Zum Beispiel Zeutschel, einer der führenden Hersteller von Geräten und Systemen für digitale und analoge Datenspeicherung, wie der Buchscanner und der Buchkopierer. [mehr..](#)

Hier bringen wir Sie auf den neuesten Stand:

DMS
07. - 09. September 2004
DMS 2004
Essen
[mehr..](#)

MEDICA
24. - 27. November 2004
MEDICA
Düsseldorf
[mehr..](#)

CeBIT
10.-16.März 2005
CeBIT 2005
Hannover
[mehr..](#)

Deutscher Bibliothekartag
2005
Düsseldorf

15.-18.März 2005
Bibliothekartag 2005
Düsseldorf
[mehr..](#)

Warum sollten Sie sich MirrorWeb ansehen?

1. bessere Technik

- Bessere Ergebnisse durch moderne Crawltechnologie: vollständige Überlieferung für das Archiv (Unterseiten, Bilder und PDFs)
- Komplexität: steigt rasant an -> „Electrolyte“
- Moderner und komfortabler Zugriff – Volltextsuche über alle Inhalte im Backend und für die Benutzer im Lesesaal/Web
- Vollständigkeit: Webarchiv inkl. aller Inhalte sowie auch Social-Media
- Standards - WARC-Format (Web ARChive)
- Sicherheit – Projektstellen, Wayback Machine, lokaler Anbieter, Marktführer

Warum sollten Sie sich MirrorWeb ansehen?

2. mehr Fokus auf Ihre Kernkompetenz

Viele Einrichtungen sammeln zu wenig und/oder können sich die Ergebnisse nicht anschauen – Sie auch?

Sie sind die Bibliothek und wir der Technikanbieter - jeder hat seine Stärken

- Mehr Zeit für Ihr Team - Automatisierung der Crawls
- Mehr Zeit für Ihr Team - Weniger QS-Aufwand, da höhere Qualität des Crawls
- Mehr Zeit für Ihr Team – keine „Basterei“ notwendig

Warum sollten Sie sich MirrorWeb ansehen?

3. Budgets senken oder mehr fürs Geld bekommen

EntscheiderInnen wissen: Open Source ist nicht umsonst!

Investitionen planbar machen

- Niedrige Einstiegskosten und nur eine Jahresgebühr für Alles
- Keine Installationsgebühren + Jederzeit kostenloser Datenexport
- Keine Kosten für eigene Hardware, Speicherung und Datensicherung
- Jederzeit kündbar
- Kosten werden rasant steigen durch Vervielfachung der Datenmenge

Beispiel: The National Archives UK

Fakten

- Archivierte Inhalte ab 1996
- Wechsel zu MirrorWeb in 2017
- ca. 130 TB Datenvolumen
- Mehr als 5.000 archivierte Webseiten
- 1.400.000.000 indexierte Inhalte/Seiten/Dokumente
- Öffentlich zugängliche Webseite:

<http://www.nationalarchives.gov.uk/webarchive/>



Beispiel: The National Archives UK

Gründe für MirrorWeb

- Geringere Kosten für Speicherung und Betrieb
- Moderne Lösung für ein komplexes Nischenthema
- „Alte Problem“ mit der Datenmenge
 - Sehr schneller Anstieg der Datenmenge
 - Langsames Handling der Daten (Index, Suche, Datensicherung etc.)
 - Datensicherheit + Redundanz der Daten (Vervielfachung des Volumens)
 - Aufwand für Qualitätssicherung
- Ergebnis heute: Mehr Zeit für andere Tätigkeiten, durch Outsourcing des Betriebs



Beispiel: The National Archives UK

LIVE



Und nun?

„Digitalisierungsoffensive 2020“- mit Karl Stringer (CPO von
MirrorWeb)

Für alle Teilnehmer bieten wir einen **kostenlosen Crawl** an.
Schreiben Sie mir.

Kontakt:

Jens Peters, j.peters@walternagel.de

<http://www.walternagel.de/webarchivierung>

<http://www.walternagel.de/social-media-archivierung>