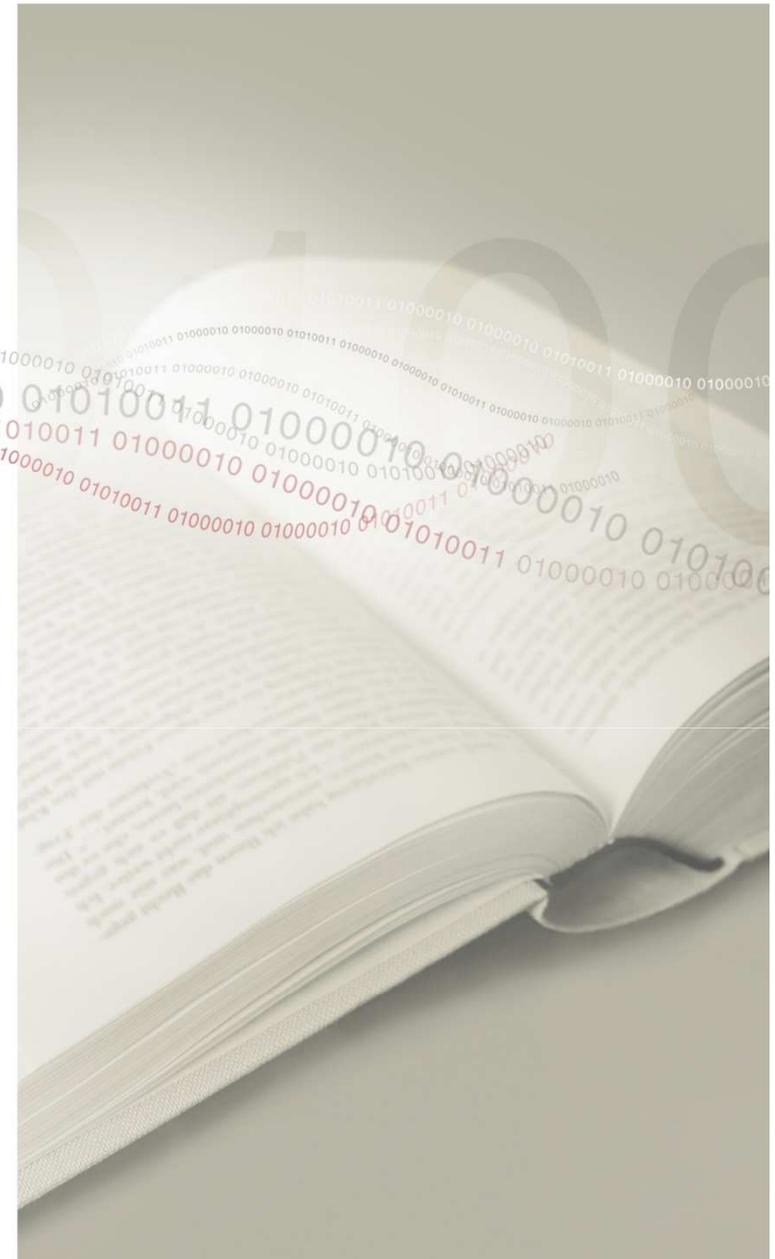


Heute ernten für die Wissenschaft von morgen

Sammlung und Archivierung von
Websites als bibliothekarische
Dienstleistung



Übersicht

- Ziele und Grundbegriffe der Webarchivierung
- Übersicht: Stand der Webarchivierung in deutschen Bibliotheken
- Webarchivierung an der Bayerischen Staatsbibliothek
- Möglichkeiten und Grenzen in der Praxis
- Entwicklungsperspektiven / Ausblick / Nutzungsszenarien

Gründe für die Webarchivierung

- Hohe Flüchtigkeit der Inhalte des Webs als Publikations-, Informations- und Kommunikationsmedium für die Wissenschaft
 - Das Internet als Ausdruck der Gegenwartskultur
 - Webarchive sind eine wichtige Quelle für Forschungsliteratur und Gegenstand kulturgeschichtlicher Forschung
- ⇒ Auffindbarkeit der Inhalte und Referenzierbarkeit bzw. Zitierbarkeit des Webs muss für die Wissenschaft dauerhaft ermöglicht werden
- ⇒ Es können immer nur Ausschnitte bzw. Schnappschüsse gesammelt werden: Ein umfassendes Bild ergibt sich erst in der Gesamtschau der unterschiedlichsten Sammlungen von Websites

Grundbegriffe der Webarchivierung

- Webarchivierung umfasst die Auswahl, Sammlung, Langzeitarchivierung, Bereitstellung von Websites bzw. Webressourcen
- Website = Webpräsenz, die aus mehreren Webpages bzw. Dokumenten bestehen kann, verknüpft durch eine logische Linkstruktur
 - ⇒ Webarchivierung meint in der Regel “Website-Archivierung”
- Harvesting / Crawling: Verfolgen der (internen) Links und lokale Speicherung der einzelnen Dateien sowie Überführung in eine Struktur bzw. Format, die den Gesamtkontext einer Website erhält
 - Domain / selektives / eventbezogenes Harvesting
 - ⇒ Entstehung von unterschiedlichen Snapshots

Webarchivierung an der DNB

- Gesetzliche Grundlage: DNB-Gesetz
- Erste Tests 2005/2007 und Mitgliedschaft im International Internet Preservation Consortium (IIPC)
- Konzeption und Ausschreibung des Verfahrens mittlerweile abgeschlossen -> derzeit: Umsetzung mit externem Dienstleister
- Auswahlkriterien: Websites von Bundesbehörden, Verbänden, gemeinnützigen Organisationen, Parteien, Kirchen / Ereignisse mit Bedeutung für Gesellschaft, Geschichte, Politik, Wirtschaft oder Kultur Deutschlands / thematische Sammlungslisten, Einzelmeldungen -> modular gestalteter Sammlungs Aufbau mit zunächst ca. 1050 Websites
- Öffentliche Zugänglichmachung: Lesesaal-Bereitstellung

Webarchivierung mit edoweb

- Betreiber seit 2003: Landesbibliothekszentrum Rheinland-Pfalz und hbz
- Gesetzliche Grundlage: Verwaltungsvorschrift für Behörden/Dienststellen + Rechteinholung für Angebote Dritter
- Lösung mit HTTrack und DigiTool inkl. Volltextsuche
- Auswahlkriterien: Websites von Landesministerien, Landesbehörden, der Kommunen sowie in Auswahl Websites wichtiger Einrichtungen des Landes und öffentlich zugängliche, privat betriebene Websites mit landeskundlichen Inhalten (ca. 500-700 Websites)
- Zugriff: Frei im Web und Lesesaal-Bereitstellung

Webarchivierungsservice SWBContent

- Anbieter seit 2002: Bibliotheksservicezentrum Baden-Württemberg
- Zahlreiche lokale Instanzen für Bibliotheken und Archive (u.a. BOA - Baden-Württembergisches Online Archiv und SaarDok)
- Technische Lösung jüngst mit Heritrix und Wayback auf den internationalen Stand gebracht
- Auswahlkriterien:
 - BOA - Websites der Landesverwaltung und mit landeskundlichen Inhalten
 - SaarDok: Institutionen des Saarlands und landeskundliche Aspekte
 - je ca. 120 Websites
- Zugriff: Frei im Web

Übersicht: Webarchivierung in dt. Bibliotheken

- Nur an wenigen Institutionen fest etablierte Geschäftsgänge für die Webarchivierung
- Aktivitäten von nestor erfreuen sich eines großen Interesses
- Fleckenteppich mit vielen regionalen und thematischen Löchern
-> Verantwortlichkeiten der Bibliotheken sind untereinander aber auch in Abgrenzung zu anderen Einrichtungen (z.B. Archiven) bislang nicht abgestimmt
- Anpassung an internationalen Entwicklungen erfolgen derzeit
- Nutzen für die Wissenschaft bislang nur partiell bzw. punktuell gegeben
-> im internationalen Vergleich ist ein Entwicklungsrückstand zu konstatieren!

Webarchivierung an der BSB

- Selektives Harvesting für die Virtuellen Fachbibliotheken
-> Sicherung der Nachhaltigkeit der Internetressourcen-Erschließung
- Harvesting der Websites der Bayerischen Staatsministerien und der Staatlichen Bibliotheken
- Halbjährliche Snapshots, derzeit ca. 150 archivierte Websites
- Genehmigungsverfahren (inkl. Rechteeinholung für Langzeitarchivierung und öffentliche Bereitstellung)
- Manuelle Qualitätskontrolle
- Zugänglichmachung erfolgt über den Katalog und die Fachportale der ViFas, Darstellung mit der Wayback-Machine

Darstellung im BSB-OPAC

OPACplus

Suche Merkliste Konto Magazin-Bestellung weitere Angebote
Einfache Suche Erweiterte Suche Suchhistorie Trefferliste Detailanzeige

Ihre Suchanfrage Freie Suche = bsbwebarch*

BSB-Katalog (10/16) |< < > >|

Archiv von: Rossijskij sojuz veteranov Afganistana (RSVA)

Ort, Jahr: S.l., archiviert ab: 2010,14.Sept. -
Sprache: Russisch

[Online lesen](#)

[In die Merkliste](#)
[Permalink](#)
[Bookmark & Share](#)

Exemplare Ausleihe / Verfügbarkeit Alle Titeldaten Weblinks

Volltext

Übersicht in der Wayback Machine

bsbwebarchiv.bsb.lrz.de:60080/wayback/wayback*/http://www.rsva.ru/  Google

Internet Adresse: 2012 [Erweiterte Suche](#)

Suche nach <http://www.rsva.ru/>

Set Anchor Window:

Suchergebnis für 01.01.2010 - 01.01.2013

2010	2011	2012
1 Seite	2 Seiten	1 Seite
14.09.2010 *	06.04.2011 * 06.10.2011 *	06.04.2012 *

[Home](#) | [Hilfe](#)

Ansicht des Archivobjekts

129.187.255.186 (www.wayback.org/wayback/20110406070025/http://www.rsva.ru/)

BSB Bayerische Staatsbibliothek
Information in erster Linie

BSB Langzeitarchiv - externe Links, Formulare und Suchabfragen werden fuer dieses Webarchiv nicht funktionieren
BSB long-term archive - external links, forms and search requests will not work for this webarchive.
Url: <http://www.rsva.ru/> - time: 06.04.2011 09:00:25 [[versteckt / hide!](#)]

ПАМЯТИ ПАВШИХ.
ВО ИМЯ ЖИВЫХ

РОССИЙСКИЙ СОЮЗ ВЕТЕРАНОВ АФГАНИСТАНА
ОБЩЕРОССИЙСКАЯ ОБЩЕСТВЕННАЯ ОРГАНИЗАЦИЯ

Трудоустройство
Медицина
Жильё

Экономика
Финансы
Бизнес

Социальная программа Законодательные акты Деловая страница Работа в регионах

Объявлен конкурс на проведение противогриппозной вакцинации членов РСВА 29.03.2011

В соответствии с решением Президиума Центрального Правления РСВА решено провести конкурс на закупку вакцины и проведение вакцинации членов РСВА.
Подробности в прилагаемых документах:
pdfcast.org/pdf/1301433849
pdfcast.org/pdf/1301434101
pdfcast.org/pdf/form-of-request [подробней »](#)

Сотрудники военкомата проигнорировали пикет афганских ветеранов 29.03.2011

<http://briansk.ru/society/sotrudniki-voenkomata-proignorirovali-piket-afganskih-veteranov.2011329.248299.html>

Бежицкое отделение Российского союза ветеранов Афганистана провело повторный пикет около Областного военного комиссариата, однако сотрудники военкомата оставили митингующих без внимания.

В пикете участвовало 22 человека. Они вышли с требованиями, среди которых – не допустить, чтобы чиновники игнорировали указания Владимира Путина, о том, что "проблемы ветеранов необходимо решать на всех уровнях власти". Участники акции потребовали предоставить списки ветеранов, умерших за послевоенное время. Списки необходимы для изучения причин высокой смертности.

Прими участие в социальной программе

Имя

Награда Президента

АРХИВ НОВОСТЕЙ

Поиск по
Апрель 2011
Пн Вт Ср Чт Пт Сб

- Open Source-Tool mit Heritrix-Crawler und Wayback Machine
- Optimiert für selektives Harvesting
- Integrierte Prozesse:
 - Genehmigungsbeantragung und -verwaltung
 - Auswahl und Scheduling von Harvest-Jobs
 - Harvesting
 - Qualitätskontrolle
 - Archivierung im WARC-Format
 - automatische Generierung deskriptiver Metadaten
- Anbindung an Discovery-Systeme der BSB und das Bibliothekarische Archivierungs- und Bereitstellungssystem (BABS)

Module des Web Curator Tool



[Home](#) | [Queue](#) | [Harvested](#) | [Help](#) | [Logout](#)
User kugler is logged in.



In Tray

26 tasks, 10 Notifications

[open](#)



Permission Request Templates

[open](#)

[add new](#)



Harvest Authorisations

226 harvest authorisations

[open](#)

[add new](#)



Reports

[open](#)



Targets

2 Targets

[open](#)



Harvester Configuration

[general](#)

[bandwidth](#)

[profile](#)



Target Instances

0 Scheduled instances, 2 ready for Quality reviews

[open](#)

[queue](#)

[harvested](#)



Users, Roles & Agencies

Users:

[open](#)

[add new](#)

Roles:

[open](#)

[add new](#)

Agencies:

[open](#)

[add new](#)



Groups

0 Target Groups

[open](#)

Herausforderungen der Praxis

- Technische Einschränkungen: Flash, Javascript, Stream Videos, Datenbanken, dynamische Inhalte nur schwer archivierbar
- > Oftmals keine vollständige Archivierung möglich/gewünscht
- > Wesentliche Eigenschaften definieren, gewisse Verluste akzeptieren!
- Fragen des Urheberrechts
- Auswahl des Archivmaterials
- Derzeit noch weitgehend manuelle Prozesse
- Abstimmung der Verantwortlichkeiten / Organisation

Perspektiven für deutsche Bibliotheken

- Bildung von thematischen Kollektionen / größeren Datensets
- Langzeitarchivierung und Langzeitverfügbarkeit
- Bessere Vernetzung von Praxis in Gedächtnisorganisationen und Forschung in der Informatik
- Fortführung und Ausbau von kooperativen Ansätzen
- Urheberrecht mitgestalten

Ausweitung der Nutzungsmöglichkeiten

- Erforschung und Darstellung der Nutzungsmöglichkeiten von Webarchiven
 - Volltextindexierung
 - Auswertung von Textcorpora mit Ngram
 - > Internet Content als Forschungsdaten
- Zielsetzung sollte ein zentraler Nutzerzugang sein
- Verbesserung der Integration in das Live-Web
 - Memento
 - Redirection Services

Fazit

- Flickenteppich mit mehreren Akteuren
- Langsam entstehendes Problembewusstsein
- 2012: Entwicklungsrückstand im internationalen Vergleich
- Hohes Entwicklungs- und Zukunftspotential für Bibliotheken und Archive
- Steigende Komplexität der Daten und deren Vernetzung sowie vielfältige neue Publikationsformen halten die Aufgabe auch in Zukunft spannend
- Es gilt nach wie vor:

Collect now - Ask later why!

Fragen?

beinert@bsb-muenchen.de

