



Mona Harring

el _ sta

Webarchivierung

Webarchivierung

Vorgeschichte...

- 2009 Erstellung des Fachkonzepts „Archivierung des Auftritts des Freistaats Sachsen im World Wide Web“
 - favorisierter Ansatz: Erfassung und Archivierung des gesamten Webauftritts des Freistaates Sachsen www.sachsen.de
- Ende 2009 Beginn der Durchführung von Tests mit dem Crawlprogramm „Offline Explorer Pro“
 - in diesem Rahmen Initiierung eines Gesamtcrawls von www.sachsen.de, der aufgrund der ausufernden Datenmenge nach einer Nacht abgebrochen werden musste

Webarchivierung

Vorgeschichte

- I Erstellung eines neuen Fachkonzepts mit folgenden Festlegungen:
 - I Erstellung ressort- bzw. behördenbezogener Crawls mit anschließender Konvertierung in ein als Archivversion fungierendes PDF-Dokument (PDF/A-Format)
 - I Archivierung erfolgt nach einem festgelegten Turnus und/oder anlassbezogen
 - I Konfiguration des Crawlerprogramms ist Ausdruck der archivfachlichen Bewertung
 - I Akzeptanz der mit der Erstellung des Crawls und dessen Weiterverarbeitung bis zur Archivvariante einhergehende Authentizitäts- und Integritätsverlust zugunsten einer optimalen Handhabbarkeit der einzelnen Archivpakete bei Archivierung und Benutzung
 - I vor jedem Archivierungsdurchgang Prüfung des für den Crawl vorgesehenen Webbereichs auf Veränderungen
 - I Dokumentation jedes durchgeführten Archivierungsdurchgangs einschließlich Vor- und Nachbereitung, hierfür Erstellung so genannter Crawlprofile

Webarchivierung

Crawlprofil (Beispiel)

Behördenbezeichnung:	Sächsisches Staatsministerium der Finanzen (SMF)	
Verwaltungsauftritt:	www.smf.sachsen.de	
nachgeordnete Behörden:	Staatsbetrieb Sächsisches Immobilien- und Baumanagement	http://www.sib.sachs.de
	Staatsbetrieb „Staatliche Schlösser, Burgen und Gärten Sachsen	http://www.schloesser.sachsen.de
	Landesamt für Finanzen	http://www.lff.sachsen.de
	Oberfinanzdirektion Chemnitz	http://www.ofd.smf.sachsen.de
	Finanzämter	http://www.finanzamt.sachsen.de
Themenportale:	Steuerportal	http://www.steuern.sachsen.de
Sonstiges:	Sächsischer Rechnungshof	http://www.rechnungshof.sachsen.de
Archivierungsintervall:	alle 11 Monate jeweils am 3. des Monats	
Projektname:	SMF + lfd. Nr. (z. B. SMF_01)	
Crawltermin:	[TT.MM.JJJJ]	

Projektname	SMF_01		
Kategorien	Details		Hinweise
Adressen (URL)	www.smf.sachsen.de	Start	
	http://www.steuern.sachsen.de	Themenportale	
	http://www.rechnungshof.sachsen.de	Sonstiges	
Stelle, von der Dateien geladen werden sollen	nur innerhalb der URL laden		
Tiefengrenze	2		
Downloads	nur modifizierte Dateien downloaden		
Dateifilter	Text		jeweils auswählen
	Bilder		
	Video		
	Audio		
	Archiv		
	Benutzerdefiniert		
	Sonstiges		„Laden mit URL-Filtereinstellungen“
URL-Filter	nicht ausfüllen	URL Ausschluss Protokoll Server	
	Alle Protokolle		
	Laden von Dateien nur aus Start-Datei		
	Server		
	Laden von Dateien nur aus dem Start-Ordner und darunter	Verzeichnis	
	Laden von Dateien nur aus dem Start-Dateinamen	Datei	
Inhalts-Filter	nicht ausfüllen		
Erweitert	nicht ausfüllen	Grenzen	
	Integrität von HTML-Dateien prüfen	Syntaxanalyse	
	Konvertiere RSS-Feed nach HTML		
	Unterdrücke Fehler auf Website		
	Offline anpassen	Linkanpassung	
	Manuell (Termin s. o.)	Zeitplaner	
	nicht ausfüllen	Passwörter	
	nicht ausfüllen	Datei Kopien	
Projektfumfang		Dateigröße	nach
		Anzahl Dateien	erfolgttem Crawl eintragen

Webarchivierung

im Rahmen des Aufbaus des el_sta

- Wechsel des Crawlerprogramms vom kostenpflichtigen „Offline Explorer Pro“ zum kostenlosen Offlinebrowser „HTTrack“
- Durchführung erster Tests mit „HTTrack“
- Modifizierung des Verfahrens:
 - Crawl wird weiterhin vom Sächsischen Staatsarchiv erstellt
 - dessen Konvertierung in ein PDF-Dokument erfolgt durch die Fa. „Scope Solutions“
 - PDF wird als dynamisches Dokument zur Verfügung gestellt:
Hyperlinks der Navigationspalte werden in Sprungmarken umgewandelt, so dass das Navigieren im Dokument weiterhin möglich ist;
Links auf Dokumente im Download-Bereich bleiben erhalten, sofern relativ zum Hauptdokument platziert

Webarchivierung

Beispiel

Crawl des Webauftritts des Sächsischen Staatsministeriums der Finanzen (SMF)

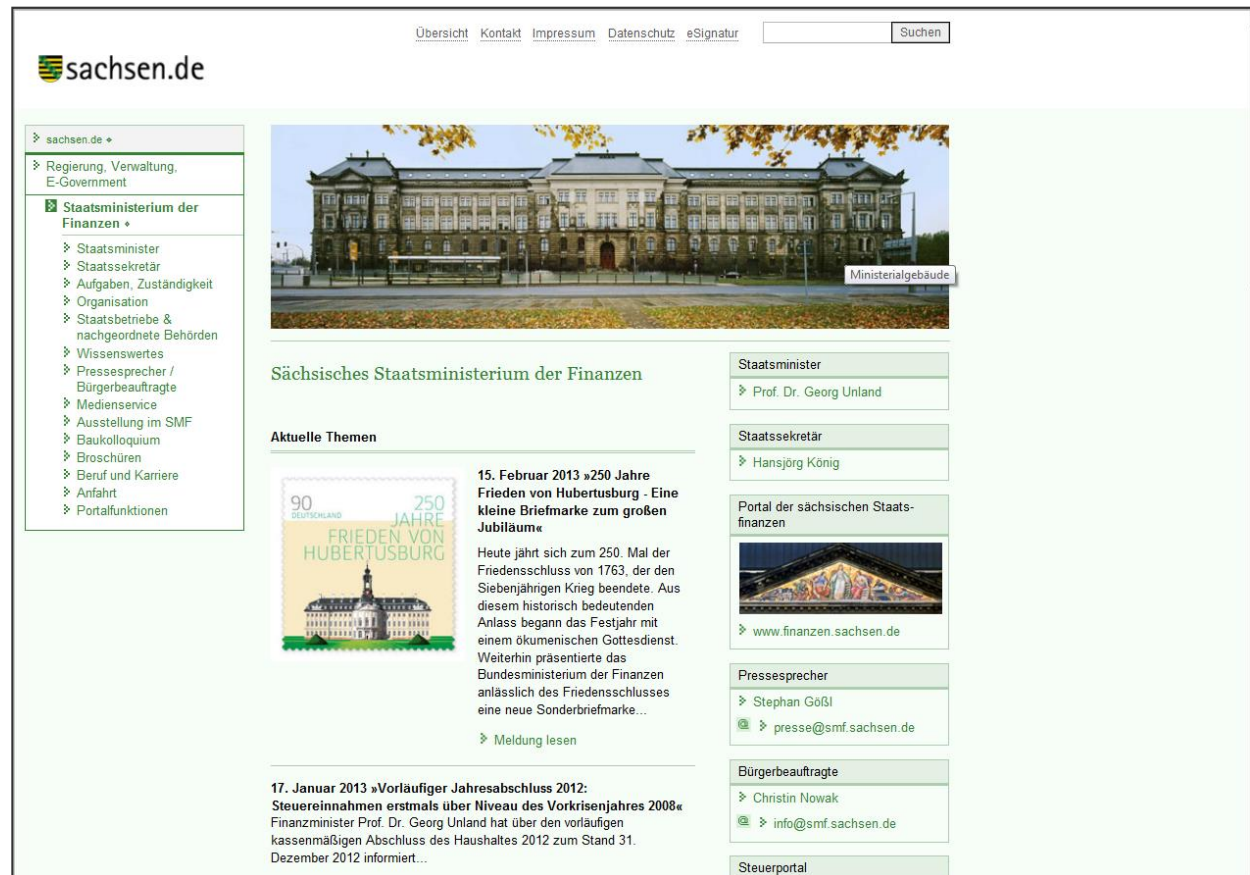
URL: <http://www.smf.sachsen.de>
(Verwaltungsportal)

Crawldatum: 8. März 2013

verwendeter Crawler: HTTrack

Crawlerkonfiguration:

- Aktion: Erstellung einer automatischen Web-Site-Kopie
- Linktiefe: intern: maximal 2; extern: maximal 0
- Filterregeln: lade +*.gif +*.jpg +*.png +*.tif +*.bmp +*.pdf
- Struktur: Verknüpfungen zu nicht kopierten externen Seiten führen zu Fehlermeldungen

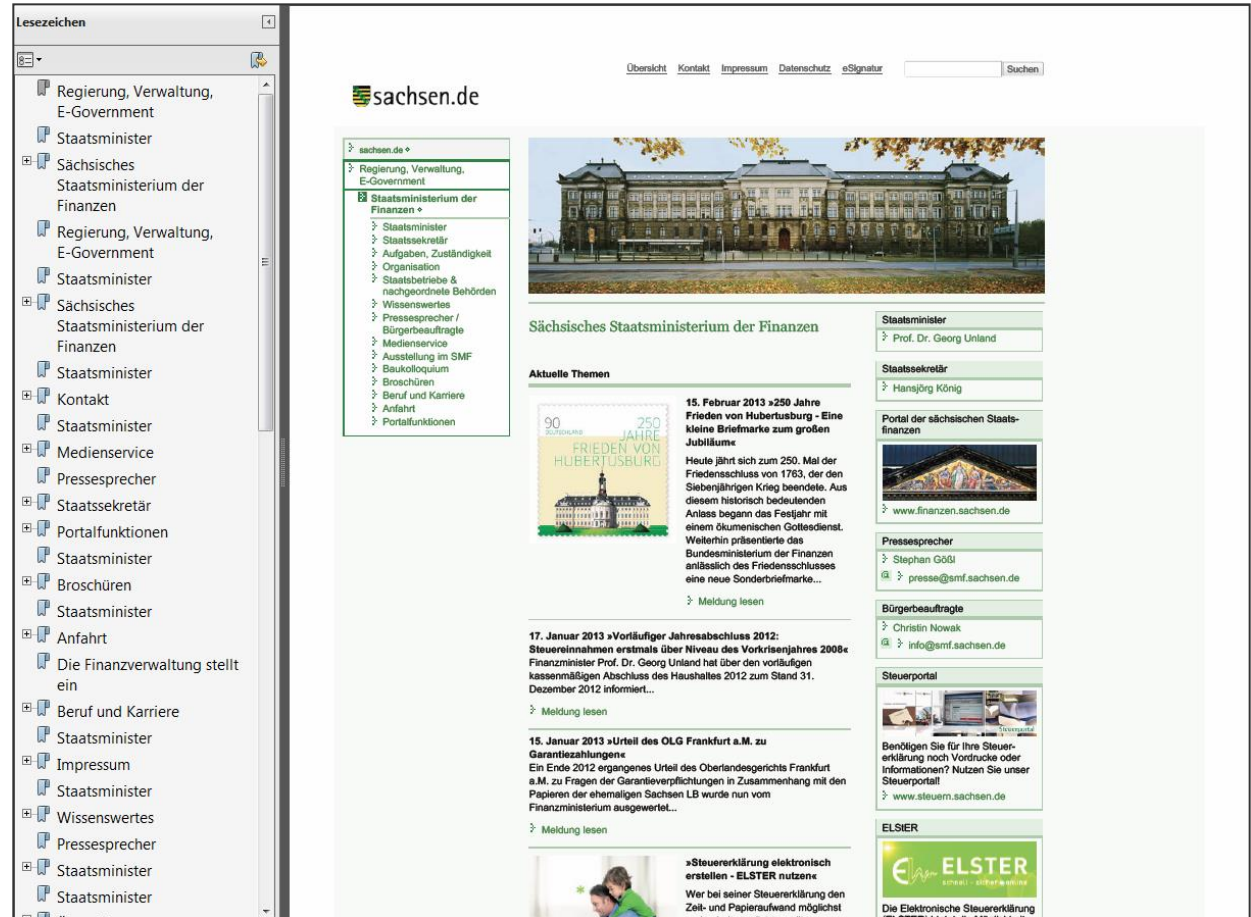


The screenshot shows the homepage of the Sächsisches Staatsministerium der Finanzen (SMF). At the top, there is a navigation bar with links for 'Übersicht', 'Kontakt', 'Impressum', 'Datenschutz', and 'eSignatur', along with a search box. The main header features the 'sachsen.de' logo and a large photograph of the 'Ministerialgebäude' in Dresden. Below the header, a left sidebar contains a menu for 'sachsen.de' with sub-items for 'Regierung, Verwaltung, E-Government' and 'Staatsministerium der Finanzen', which is expanded to show various internal and external links. The main content area is titled 'Sächsisches Staatsministerium der Finanzen' and includes a section for 'Aktuelle Themen' with a featured article about the 250th anniversary of the Peace of Hubertusburg. On the right side, there are several vertical panels for 'Staatsminister' (Prof. Dr. Georg Unland), 'Staatssekretär' (Hansjörg König), 'Portal der sächsischen Staatsfinanzen', 'Pressesprecher' (Stephan Gößl), 'Bürgerbeauftragte' (Christin Nowak), and 'Steuerportal'.

Webarchivierung

Beispiel

aus dem Crawl des
Webauftritts des SMF
erzeugtes PDF-Dokument
(mit aktiviertem Lesezeichen)



The image shows a browser window with a bookmarked page. On the left, a sidebar titled 'Lesezeichen' (Bookmarks) lists various categories and sub-items, including 'Regierung, Verwaltung, E-Government', 'Staatsminister', 'Sächsisches Staatsministerium der Finanzen', 'Regierung, Verwaltung, E-Government', 'Staatsminister', 'Sächsisches Staatsministerium der Finanzen', 'Staatsminister', 'Kontakt', 'Staatsminister', 'Medienservice', 'Pressesprecher', 'Staatssekretär', 'Portalfunktionen', 'Staatsminister', 'Broschüren', 'Staatsminister', 'Anfahrt', 'Die Finanzverwaltung stellt ein', 'Beruf und Karriere', 'Staatsminister', 'Impressum', 'Staatsminister', 'Wissenswertes', 'Pressesprecher', 'Staatsminister', and 'Staatsminister'. The main content area displays the website 'sachsen.de' for the 'Sächsisches Staatsministerium der Finanzen'. The page features a header with navigation links (Übersicht, Kontakt, Impressum, Datenschutz, eSignatur) and a search bar. Below the header is a large image of the ministry building. The main content includes 'Aktuelle Themen' with several news items: '15. Februar 2013 »250 Jahre Frieden von Hubertusburg - Eine kleine Briefmarke zum großen Jubiläum', '17. Januar 2013 »Vorläufiger Jahresabschluss 2012: Steuereinnahmen erstmals über Niveau des Vorkrisenjahres 2008', and '15. Januar 2013 »Urteil des OLG Frankfurt a.M. zu Garantiezahlungen'. On the right side, there are contact information sections for 'Staatsminister' (Prof. Dr. Georg Untland), 'Staatssekretär' (Hansjörg König), 'Pressesprecher' (Stephan Gößl), and 'Bürgerbeauftragte' (Christin Nowak). At the bottom right, there is a section for 'ELSTER' (Elektronische Steuererklärung).

Webarchivierung

Was bleibt noch zu tun

- Vor Überführung in den Produktivbetrieb:
 - Durchführung weiterer Tests mit HTTrack und PDF-Konvertierung v. a. aus Gründen der Optimierung von Verfahren und Ergebnis
 - Schulung der für die Webarchivierung verantwortlichen Mitarbeiter
 - Entwicklung von Lösungen zur Erschließung und Bereitstellung / Benutzung