UB Basel. Informatik

Aufbereiten der hierarchischen Webpräsentation der   
SWA-Archivbestände

Technische Dokumentation / Andres von Arx

Inhalt

[Sachdokumentation 2](#_Toc455042011)

[Präsentation im Web 2](#_Toc455042012)

[Generieren der Webseiten 2](#_Toc455042013)

[Datenfluss 3](#_Toc455042014)

[Firmen- und Verbanddokumentation 7](#_Toc455042015)

[Präsentation im Web 7](#_Toc455042016)

[Generieren der Webseiten 7](#_Toc455042017)

[Datenfluss 8](#_Toc455042018)

[Privatnachlässe 10](#_Toc455042019)

[Präsentation im Web 10](#_Toc455042020)

[Generieren der Webseiten 10](#_Toc455042021)

[Systematische Gliederung der E-Zeitungsausschnitte (E-ZAS) 11](#_Toc455042022)

[Präsentation im Web 11](#_Toc455042023)

[Generieren der Webseiten 11](#_Toc455042024)

[Datenfluss 12](#_Toc455042025)

# Sachdokumentation

Die Deskriptoren der Sachdokumentation basieren auf einer lokalen, noch unverlinkten Version des „Standard-Thesaurus Wirtschaft“. Die Webseiten sind ziemlich stabil. Ein Wechsel auf den STW (linked Version) ist angedacht. Der aktuelle Thesaurus ist als Aleph Autoritätsdatei implementiert. Vom Einrichten einer neuen Aleph Autoritätsdatei für den STW wird abgeraten, da kaum nachhaltig.

## Präsentation im Web

<http://www.ub.unibas.ch/ub-wirtschaft-swa/schweiz-wirtschaftsarchiv/sachdokumentation/>

## Generieren der Webseiten

Anleitung:

<http://ub-files.ub.unibas.ch/ordnungssystem/1_Eigene-Organisation/15_IT/153_Bibliothekssysteme/bibtools/hierarchie_toolbox/swa_sachdokumentation/ANLE_SWA_Sachdokumentation_aufbereiten_20140818_ava.html>

Wie: Aleph GUI  
Linux-Client mit Zugriff auf ub-files und SSH public key auf ub-webqm/ub-webvm

Wann: bei Bedarf, auf Bitte vom SWA

Laufzeit: Aufbereitung der Aleph-Daten: 5-10 Minuten  
Konversionsskript: 3 Minuten

$ cd ~/intranet/ordnungssystem/1\_Eigene-Organisation/15\_IT/153\_Bibliothekssysteme

$ cd bibtools/hierarchie\_toolbox/swa\_sachdokumentation

$ ./make-swa-sachdok.sh

Generiere Webseiten für die SWA Sachdokumentation

-------------------------------------------------

\* Dateien von Aleph herunterladen

Anzahl Aleph Records:

BIB: 1267

AUT: 963

\* flicke Zeichensatz

\* konvertiere nach MARC21 XML

\* extrahiere relevante AUT + BIB Daten

\* kombiniere AUT + BIB zu Strukturdatei

\* extrahiere Hierarchie als Textdatei

\* extrahiere Hierarchie (Firmen+Verbände)

\* dokumentiere Thesaurus

\* generiere Listen für E-Zeitschriftenarchiv

\* generiere Tree Menu

\* generiere Detailseiten

\* generiere Detailseiten (Firmen+Verbände, ohne Geographica)

\* generiere A-Z-Liste

\* generiere Wortindex für Suchmaschine

\* generiere A-Z-Liste für Suchmaschine (Firmen+Verbände)

\* generiere Wortindex für Suchmaschine (Firmen+Verbände)

\* Outputdateien zippen und hochladen

fertig: 24. Juni 2016, 12:25:16

## Datenfluss

(vereinfacht, Zwischendateien sind ausgelassen)

aleph.unibas.ch

Bibliographische Datei  
DSV01

Autoritätsdatei  
DSV12

Selektion: wlc=doksf

Selektion: wlc=swa

pswabib.seq

pswaaut.seq

ub-nx

htb\_alephseq\_to\_marcxml

htb\_alephseq\_to\_marcxml

aut-extract.xsl

bib-extract.xsl

swasd-hierarchie.pl

swasd-hierarchie.xml

Die Autoritätsdatei enthält Aufnahmen für jeden Deskriptor, mit impliziter Hierarchie (über­geord­nete und verwandte Begriffe, Synonyme). Die Bibliographische Datei enthällt Aufnahmen für die einzelnen Dossiers mit den entsprechenden Deskriptoren; diese können ev. noch geographischen Unterteilungen enthalten.

Die Programme generieren aus diesen Daten eine hierarchisch gegliederte Datei **swasd-hierarchie.xml**, die alle Informationen zusammenfasst.

Diese Datei ist Ausgangspunkt der folgenden Produkte:

swasd-descriptors.xsl

../htb/generate\_treemenu\_pro.pl

Sachdokumentation:  
Steuerdatei von Tigra Tree Menu.

Javascript für Hierarchiebaum.

output/tree\_items.js

Sachdokumentation: rund 1300 Seiten für einzelne Deskriptoren, mit Synonymen und verwandten Begriffen, Vorkommen des Deskiptors in anderem Kontext und Deep-Linking zu Aleph.

swasd-detailseiten.xsl

swasd-splitte-detailseiten.pl

output/details/did\*.html

swasd-indexterms.xsl

swasd-index-az.xsl

swasd-index-az.pl

Sachdokumentation:  
Indexseiten A-Z

output/index\*.html

swasd-indexterms.xsl

swasd-suchmaschine.xsl

swasd-suchmaschine.pl

Sachdokumentation:  
Wortliste für eine einfache sequentielle Volltextsuche mit  
 /cgi-bin/index/swasd-index.pl

tmp/suchmaschine.txt

Der Teilbaum „Wirtschaftssektoren“ der Sachdokumentation ist gleichzeitig die Systematische Gliederung der Firmen- und Verbanddokumentation (s.u.). Die dafür benötigten Dateien werden ebenfalls hier aufgebaut und für die Weiterbearbeitung in den Skripten für die F+V Doku bereitgestellt[[1]](#footnote-1).

swafv-descriptors.xsl

Firmen- + Verbanddokumentation:

Grundgerüst des Hierarchiebaums.

swafv/swafv-descriptors.txt

Firmen- + Verbanddokumentation:

Texte zu den einzelnen Deskriptoren, noch ohne konkrete Körperschaften, ungesplittet.

swafv-detailseiten.xsl

swafv/detailseiten.html

swafv-indexterms.xsl

swasd-index-az.xsl

Firmen- + Verbanddokumentation: Rohdaten für die A-Z Listen

swafv/indexterms.txt

swafv/suchmaschine\_deskriptoren.tmp

swasd-suchmaschine.xsl

swafv-indexterms.xsl

Firmen- + Verbanddokumentation: Rohdaten für die Volltextsuche

Die Deskriptoren der Sachdokumentation bilden auch einen Bestandteil des Thesaurus des Elektronischen Zeitschriften-Ausschnitt-Dienstes (E-ZAS, s.u.). Die dazu benötigten Dateien werden ebenfalls hier generiert.

swafv-descriptors.xsl

E-ZAS:  
Baum der Sach-Deskriptoren (Text)

doku-ezas/swa-sachdokumentation-deskriptoren.txt

swasd-descriptors-doku.xsl

E-ZAS:  
Baum der Sach-Deskriptoren (XML)

doku-ezas/swa-sachdokumentation-deskriptoren.xml

swasd-e-zas-flach.xsl

doku-ezas/swa-e-zas-sachthemen-flach-unicode.txt

E-ZAS:  
Baum der Sach-Deskriptoren (Text) mit Geographica

doku-ezas/swa-e-zas-sachthemen-flach-windows.txt

E-ZAS:  
Deskriptoren Sachdokumentation mit Geographika, 5-stufige Ketten für den Import in NewBase/E-ZAS

doku-ezas/swa-e-zas-sachthemen-hierarchie-unicode

swasd-e-zas-hierarchie.xsl

# Firmen- und Verbanddokumentation

## Präsentation im Web

<http://www.ub.unibas.ch/ub-wirtschaft-swa/schweiz-wirtschaftsarchiv/firmen-und-verbandsdokumentation/>

## Generieren der Webseiten

Anleitung:

<http://ub-files.ub.unibas.ch/ordnungssystem/1_Eigene-Organisation/15_IT/153_Bibliothekssysteme/bibtools/hierarchie_toolbox/swa_firmen_und_verbaende/ANLE_SWA_Firmen_und_Verbaende_aufbereiten_20140604_ava.html>

Wie: Linux-Client mit Zugriff auf ub-files und SSH public key auf ub-webqm/ub-webvm

Wann: Jeden Freitag morgen

Laufzeit: 5-10 Minuten

$ cd ~/intranet/ordnungssystem/1\_Eigene-Organisation/15\_IT/153\_Bibliothekssysteme

$ cd bibtools/hierarchie\_toolbox/swa\_firmen\_und\_verbaende

$ ./make-firmen-und-verbaende.sh

SWA Firmen und Verbände: Webseiten aufbauen

-------------------------------------------

Fre Jun 24 15:00:57 CEST 2016

\* Aleph Daten downloaden

Daten vom 23 Jun 2016 22:05:04

\* Qualitäts-Check

\* Konkordanz Deskriptoren / Körperschaften

\* Extrahiere Körperschaften

\* Geographische Unterschlagwörter extrahieren

\* Branchen-Systematik mit Geographica anreichern

[Anmerkung: nur Geographica, die in den doksf-Daten vorkommen.]

\* Tree Menu generieren

\* Detailseiten aufbereiten

(bitte 5 Minuten Geduld...)

\* generiere Wortindex (Deskriptoren)

\* generiere Wortindex (Körperschaften)

\* generiere A-Z Liste

\* Outputdateien zippen und hochladen

\* installiere Webseiten auf Webserver (TEST)

\* installiere Wortindex (Deskriptoren) auf Webserver (TEST)

\* installiere Wortindex (Körperschaften) auf Webserver (TEST)

\* installiere Webseiten auf Webserver (PROD)

\* installiere Wortindex (Deskriptoren) auf Webserver (PROD)

\* installiere Wortindex (Körperschaften) auf Webserver (PROD)

\* temporäre Dateien löschen

Alle Körperschaften sind GND-kontrolliert.

Fre Jun 24 15:02:08 CEST 2016

fertig.

## Datenfluss

(vereinfacht, Zwischendateien sind ausgelassen)

Bibliographische Datei  
DSV01

Selektion: wlc=doksf

swa-doksf.seq

aleph.unibas.ch ub-nx ../from\_swad  
 [zusammen mit Sachdok aufbereitet]

1

indexterms.txt

2

suchmaschine\_deskriptoren.tmp

detailseiten.html

descriptors.txt

extract-firmen-deskriptoren.pl

kss-desk.txt

add-geographica-to-descriptors.pl

desk-geographica.txt

kss-aut.txt

3

swafv-descriptors-plus.txt

geographica.tmp

../htb/generate\_treemenu\_pro.pl

Firmen + Verbände:  
Steuerdatei von Tigra Tree Menu.

Javascript für Hierarchiebaum.

output/tree\_items.js

swafv-splitte-detailseiten.pl

Firmen + Verbände:  
rund 655 Detailseiten der „Wirtschafts­sektoren“ mit Kontext und Liste der zugehörigen Firmen und Verbände.

output/details/did\*.html

swafv-index-az.pl

1

Firmen + Verbände:  
Indexseiten (Deskriptoren) A-Z

output/index\*.html

suchmaschine-swafv-desk.pl

2

Firmen + Verbände: Wortliste für Volltextsuche (Deskriptoren)

tmp/index\_swafv\_desk

tmp/index\_swafv\_kss

Firmen + Verbände: Wortliste für Volltextsuche (Körperschaften)

suchmaschine-swafv-kss.pl

3

# Privatnachlässe

## Präsentation im Web

<http://www.ub.unibas.ch/ub-wirtschaft-swa/schweiz-wirtschaftsarchiv/privatarchive/>

## Generieren der Webseiten

Anleitung:

<http://ub-files.ub.unibas.ch/ordnungssystem/1_Eigene-Organisation/15_IT/153_Bibliothekssysteme/bibtools/hierarchie_toolbox/swa_privatarchive/ANL_SWA_Privatarchive_Daten_generieren_20130617_ava.html>

Wie: Linux-Client mit Zugriff auf ub-files und SSH public key auf ub-webqm/ub-webvm

Wann: bei Bedarf, manuell

Laufzeit: 5 Minuten (Download von DSV05)  
 1 Minute (Laufzeit Skript)

$ cd ~/intranet/ordnungssystem/1\_Eigene-Organisation/15\_IT/153\_Bibliothekssysteme

$ cd bibtools/hierarchie\_toolbox/swa\_privatarchive

$ ./make-swapa.sh

SWA Privatarchive

-----------------

\* extrahiere Daten aus Aleph Sequential (dsv05)

\* konvertiere in MARC XML

\* generiere Schlagwort-Struktur

\* generiere Hierarchie

\* generiere Tree Menu

\* fixe Tree Menu

\* baue Index für Suchmaschine

\* installiere Treemenu auf Webserver

\* installiere Wortindex auf Webserver

\* cleanup tmp files

# Systematische Gliederung der E-Zeitungsausschnitte (E-ZAS)

## Präsentation im Web

<http://ezas.wirtschaftsarchiv.ch/>

## Generieren der Webseiten

Anleitung:

<http://ub-files.ub.unibas.ch/ordnungssystem/1_Eigene-Organisation/15_IT/153_Bibliothekssysteme/E-ZAS/thesaurus/ANLE_ezas_thesaurus_aufbereiten_20140501_ava.html>

$ cd ~/intranet/ordnungssystem/1\_Eigene-Organisation/15\_IT/153\_Bibliothekssysteme

$ cd E-ZAS/thesaurus/bin

$ ./make-ezas-thesaurus.sh

-------------------------------------

Generiere E-ZAS Thesaurus

-------------------------------------

\* download Aleph-Dateien

\* pruefe Pfade und Dateien

\* Sachthemen

\* Firmen und Verbände

\* Personen

\* kombiniere Teile

Thesaurus.sdf

Thesaurus\_kurz.sdf

\* dokumentiere Unterschiede

DOKU\_Thesaurus\_Aenderungen.html

DOKU\_Thesaurus\_Aenderungen\_komplett.html

\* cleanup

done.

## Datenfluss

(vereinfacht, ohne Zwischenschritte)

Der Systematische Gliederung in E-ZAS speist sich aus verschiedenen Quellen:  
- Firmen und Verbände: Körperschaftsansetzungen aus Aleph Dossieraufnahmen („doksf“)

- Personen: Lokal- und GND-Personenschlagwörter aus Aleph Dossieraufnahmen („doksb“)

- Sachthemen: Deskriptoren aus der Sachdokumentation (s.o.)

aleph.unibas.ch ub-nx

swa\_sachdokumentation/doku-ezas

Bibliographische Datei  
DSV01

Selektion: wlc=doksb / 600##

swa-doksb-600

Bibliographische Datei  
DSV01

<

[aufbereitet zusammen mit

der Sachdokumentation.]

Selektion: wlc=doksf / 710##

swa-doksf-71x

swa-e-zas-sachthemen-hierarchie

ezas-thesaurus-extract-600.pl

ezas-thesaurus-extract-71x.pl

ezas-thesaurus-personen.pl

ezas-thesaurus-firmen.pl

PERSONEN.\*

FIRMEN.\*

Thesaurus.sdf

Die Datei Thesaurus.sdf bildet die Systematische Gliederung innerhalb von E-ZAS.   
Sie wird vom SWA direkt in E-ZAS eingespielt.

Die Programme generieren darüber hinaus eine Dokumentation der Änderungen in Thesaurus.sdf. Mit Hilfe dieser Dokumentation können allenfalls die Deskriptoren in E-ZAS manuell geändert werden.

ACHTUNG: In der Nachfolge der Einführung von GND sind grössere Bereinigungsarbeiten an den Deskriptoren nötig. Das SWA (gfr) ist in Kontakt mit der Herstellerfirma Newbase, um sich Tools zur Batch-Mutation von Deskriptoren offerieren zu lassen. Eine Konkordanzliste alt/neu existiert. (Stand Juni 2016/ava)

1. **Achtung**: diese beiden GEVER Verzeichnis sind ein und dasselbe (Unix hard link):  
   …/153\_Bibliothekssysteme/bibtools/hierarchie\_toolbox/swa\_sachdokumentation/**swafv**  
   …/153\_Bibliothekssysteme/bibtools/hierarchie\_toolbox/swa\_firmen\_und\_verbaende/**from\_swasd** [↑](#footnote-ref-1)