

H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, M. Jehn (Hrsg.)

nestor Handbuch

Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung

Version 2.0

Kapitel 13.4
Die kopal Library for
Retrieval and Ingest (koLibRI)

nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung
hg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, M. Jehn
im Rahmen des Projektes: nestor – Kompetenznetzwerk Langzeitarchivierung und
Langzeitverfügbarkeit digitaler Ressourcen für Deutschland
nestor – Network of Expertise in Long-Term Storage of Digital Resources
<http://www.langzeitarchivierung.de/>

Kontakt: editors@langzeitarchivierung.de
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen,
Dr. Heike Neuroth, Forschung und Entwicklung, Papendiek 14, 37073 Göttingen

Die Herausgeber danken Anke Herr (Korrektur), Martina Kerzel (Bildbearbeitung) und
Jörn Tietgen (Layout und Formatierung des Gesamttextes) für ihre unverzichtbare
Unterstützung bei der Fertigstellung des Handbuchs.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter
<http://www.d-nb.de/> abrufbar.

Die Inhalte dieses Buchs stehen auch als Onlineversion
(<http://nestor.sub.uni-goettingen.de/handbuch/>)
sowie über den Göttinger Universitätskatalog (<http://www.sub.uni-goettingen.de>) zur
Verfügung.

Die digitale Version 2.0 steht unter folgender Creative-Commons-Lizenz:
„Attribution-Noncommercial-Share Alike 3.0 Unported“
<http://creativecommons.org/licenses/by-nc-sa/3.0/>



Einfache Nutzungsrechte liegen beim Verlag Werner Hülsbusch, Boizenburg.
© Verlag Werner Hülsbusch, Boizenburg, 2009
www.vwh-verlag.de
In Kooperation mit dem Universitätsverlag Göttingen

Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen,
Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und
als solche den gesetzlichen Bestimmungen unterliegen.

Druck und Bindung: Kunsthaus Schwanheide

Printed in Germany – Als Typoskript gedruckt –

ISBN: 978-3-940317-48-3

URL für Kapitel 13.4 „Die kopal Library for Retrieval and Ingest (koLibRi)“ (Version 2.0):
<urn:nbn:de:0008-20090811707>
<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20090811707>



*Gewidmet der Erinnerung an Hans Liegmann (†), der als Mitinitiator und früherer
Herausgeber des Handbuchs ganz wesentlich an dessen Entstehung beteiligt war.*

13.4 Die kopal Library for Retrieval and Ingest (koLibRI)

Stefan E. Funk

Einführung

Die *kopal Library for Retrieval and Ingest* ist ein Framework zur Integration eines Langzeitarchivs wie dem IBM Digital Information Archiving System²⁰ (DIAS) in die Infrastruktur einer Institution. Insbesondere organisiert koLibRI das Erstellen und Einspielen von Archivpaketen in DIAS und stellt Funktionen zur Verfügung, um diese abzurufen und zu verwalten. koLibRI stellt eine Bibliothek von Java-Tools dar, die im Projekt kopal entwickelt wurden. Sie wurde bewusst so angelegt, dass sie als Ganzes oder in Teilen auch in anderen Zusammenhängen nachnutzbar ist.

An dieser Stelle soll der Teil koLibRIs beschrieben und beispielhaft dargestellt werden, der für das Erstellen von Archivpaketen verantwortlich ist; eine ausführliche Beschreibung der gesamten Funktionalität der kopal Library for Retrieval and Ingest sowie weitere technische Details ist in deren Dokumentation²¹ zu finden und auch auf der Internetseite des Projekts kopal²².

Funktionsweise

Im einfachen Fall generiert koLibRI aus den mit dem zu archivierenden Objekt gelieferten Metadaten sowie den von JHOVE²³ maschinell extrahierten Metadaten eine XML-Datei nach METS Schema, verpackt diese zusammen mit dem Objekt in einer Archivdatei (.zip oder .tar) und liefert diese Datei als Submission Information Package (SIP) an das DIAS.

So gesehen ist koLibRI für eine vollständige Langzeitarchivierungslösung mit DIAS entwickelt worden. Jedoch kann koLibRI auch als eigenständige Software zur Generierung der METS Dateien oder kompletten SIPs nach dem Universellen Objektformat²⁴ vollkommen ohne DIAS eingesetzt werden. Die auf diese Weise generierten XML-Metadateien oder die kompletten SIPs kön-

20 <http://www-5.ibm.com/nl/dias/>

21 http://kopal.langzeitarchivierung.de/kolibri/koLibRI_v1_0_dokumentation.pdf

22 <http://kopal.langzeitarchivierung.de/>

23 JSTOR/Harvard Object Validation Environment (JHOVE):
<http://hul.harvard.edu/jhove/>

24 http://kopal.langzeitarchivierung.de/downloads/kopal_Universelles_Objektformat.pdf

nen für den Datenaustausch zwischen verschiedenen Institutionen verwendet werden; ein Aspekt, der bei der Entwicklung des UOF besonders im Vordergrund stand. Alternativ kann durch den modularen Aufbau der koLibRI auch mit vertretbarem Aufwand an ein anderes Archivsystem oder ein anderes Metadatenformat angepasst werden, da die Schnittstellen ausreichend spezifiziert sind.

Mit koLibRI kann ein für das Erstellen der Archivpakete benötigter Workflow abgebildet werden. Dieser Workflow kann den eigenen Bedürfnissen angepasst und erweitert werden. Die koLibRI-Infrastruktur nutzt prinzipiell vier Konstrukte, um Workflows abzubilden und zu verarbeiten:

- Zunächst sammelt der sogenannte *ProcessStarter* die einzuspielenden Dateien/Daten ein, die als kleinste Einheit definiert wurden – in unserem Beispiel wird dies ein Verzeichnis mit beliebigen Dateien sein.
- Jede Einheit wird vom *ProcessStarter* an die *ProcessQueue* angehängt, die dann der Reihe nach (oder auch nebenläufig) abgearbeitet werden.
- In den *ActionModules* werden einzelne Aufgaben implementiert, die für ein jedes Objekt in der *ProcessQueue* durchgeführt werden sollen. Für den Beispiel-Workflow werden hier folgende Module genutzt: *FileCopyBase*, *MetadataExtractorDmd*, *MetadataGenerator*, *MetsBuilder* und *Zip*. Weitere Module werden mit koLibRI geliefert und können integriert werden.
- Die Reihenfolge, in der die *ActionModules* für jedes dieser Objekte in der *ProcessQueue* verarbeitet werden, wird als *Policy* konfiguriert.

Installation und Konfiguration

Download

Zunächst wird das koLibRI-Paket von der kopal-Homepage benötigt, bitte laden Sie diese von der Internetseite des Projekts kopal. Benötigt wird das gepackte Programmpaket „kopal Library for Retrieval and Ingest“ (http://kopal.langzeitarchivierung.de/kolibri/koLibRI_v1_0.zip), das Sie bitte in ein beliebiges Verzeichnis entpacken.

Anforderungen

Da die kopal Library for Retrieval and Ingest komplett in Java implementiert wurde, sollte die Software prinzipiell auf jeder Plattform laufen, die eine Java Virtual Machine in der Version 1.5 zur Verfügung stellt. Alle weiteren

erforderlichen Java Software-Bibliotheken sind in dem Paket enthalten und vorkonfiguriert.

Konfiguration des Workflowtool Skriptes

Zunächst müssen die folgenden Werte in den beiden Startskripten `workflow-tool` (Linux/Unix) oder `workflowtool.bat` (Windows) an die lokalen Verhältnisse angepasst werden:

- `KOLIBRI_HOME`

Hier tragen Sie bitte den Pfad zu Ihrer koLibRI-Installation ein, zum Beispiel `/home/funk/kolibri_v1_0` (Linux/Unix) bzw. `C:\Programme\kolibri_v1_0` (Windows).

- `JAVA_HOME`

Sollte hier der Pfad zu Ihrer Java-Installation nicht stimmen, passen Sie diesen bitte ebenfalls noch an.

Konfiguration der Policies-Datei

Die Datei `policies.xml` im Verzeichnis `config/` wird um die folgenden Zeilen ergänzt; vor dem letzten schließenden Tag `</policies>` – fügen Sie bitte die folgenden Zeilen ein:

```
<policy name="example_lza_handbuch">
  <step class="FileCopyBase">
    <step class="XorFileChecksums">
      <step class="MetadataExtractorDmd">
        <step class="MetadataGenerator">
          <step class="MetsBuilder">
            <step class="Zip">
              <step class="CleanPathToContentFiles"/>
            </step>
          </step>
        </step>
      </step>
    </step>
  </step>
</policy>
```

Konfiguration der Konfigurations-Datei

In der Datei config.xml im Verzeichnis config/ werden folgende Werte gesetzt:

- Der Wert der Eigenschaft des Feldes <field>defaultPolicyName</field> wird in den Wert <value>example_lza_handbuch</value> geändert, so wird unsere vorher hinzugefügte Policy genutzt.
- Die Werte der Felder logfileDir, destinationDir, workDir und tempDir werden jeweils mit dem Pfad zu den jeweiligen Verzeichnissen ersetzt. Bitte legen Sie diese vorher an, am besten direkt in Ihren kolibri-Verzeichnis (Beispielsweise als „log“, „dest“, „work“ und „temp“): <value>./log/</value>, <value>./dest/</value>, <value>./work/</value> und <value>./temp/</value>.
- Schließlich wird noch ein Verzeichnis als Hotfolder benötigt, aus dem die zu behandelnden Dateien hineinkopiert werden. Bitte legen Sie ein weiteres Verzeichnis „./hotfolder“ an, dessen Wert sollte bereits in der Konfigurations-Datei eingetragen sein.

Starten von koLibRI

Zum Starten des Workflowtools wechseln Sie bitte in das Verzeichnis der koLibRI-Installation – oder bleiben gleich dort, sollten Sie schon da sein – und tippen

```
./workflowtool -c config/config.xml (Linux/Unix)
```

bzw.

```
workflowtool /c config\config.xml (Windows)
```

Sie bekommen nun – wenn nun alles richtig konfiguriert ist – eine Ausgabe auf der Konsole, die mit den folgenden Zeilen endet:

```
[INFO]          Checking hotfolder /Users/fugu/Desktop/koLibRI_v1_0/./hotfolder for new content
```

```
[INFO]
```

```
All current files scheduled, waiting for more
```

Nun können Sie testweise ein beliebiges Verzeichnis in dieses Hotfolder kopieren (bitte zu Anfang mit nicht allzuviel Inhalt!), und koLibRI fängt an zu arbeiten.

Ergebnis

Als Ergebnis erhalten Sie im Verzeichnis `dest` eine `.zip`-Datei, in der sich zum einen Ihre im Hotfolder befindlichen Dateien befinden und außerdem eine METS-Datei im Universellen Objektformat mit dem Namen `mets.xml`. Diese enthält neben den in der Template-XML-Datei `config/uof_template.xml` enthaltenen Daten – die für die METS-Datei als Vorlage genommen wird – technische Metadaten für jede einzelne Datei (extrahiert von JHOVE) im LMERfile-Format, sowie Metadaten zu dem gesamten Objekt im LMERobject-Format²⁵.

Weitere Konfigurationsmöglichkeiten der `kopal Library for Retrieval and Ingest` – von denen es noch viele gibt – sowie eine ausführliche Beschreibung der Nutzung auch mit dem DIAS, und weitere Nutzungsszenarien und Erweiterungsmöglichkeiten der `koLibRI` sind in der ausführlichen Dokumentation nachzulesen. Weiterhin gibt es die Möglichkeit, über die `koLibRI`-Internetseite den Entwicklern Rückmeldungen zu Erfahrungen mit `koLibRI` mitzuteilen.

Literatur

Funk, Stefan; Kadir Karaca Koçer, Sabine Liess, Jens Ludwig, Matthias Neubauer: *kopal Library for Retrieval and Ingest – Dokumentation* –. 2007. http://kopal.langzeitarchivierung.de/kolibri/koLibRI_v1_0_dokumentation.pdf

25 <http://www.d-nb.de/standards/lmer/lmer.htm>