

MABLe+ Automatische Fehleranalyse

MABLe+ beantwortet die folgenden Fragen:

- ◆ Wie oft ist welches Feld vorhanden?
- ◆ Wie hoch ist die Sacherschließung in meinem Katalog?
- ◆ Sind die Erschließungsdaten formal korrekt?
- ◆ Sind alle Pflichtfelder vorhanden?
- ◆ Sind alle Referenzen (Überordnungen etc.) vorhanden?
- ◆ Gibt es unerlaubte Felder?
- ◆ Ist der Zeichensatz korrekt?

Jeder Katalog enthält fehlerhafte Daten, die das Auffinden von vorhandenen Informationsmedien erschweren und im schlimmsten Fall unmöglich machen.

Dieses Problems haben sich der Bibliotheksverbund Bayern (BVB) und der Kooperative Bibliotheksverbund Berlin Brandenburg (KOBV) mit der Software MABLe+ angenommen.

Mit MABLe+ entsteht eine offene Software, die Katalogdaten formal prüft, statistische Auswertungen über die Verteilung von Feldern ausgibt und letztendlich automatisierte Bereinigungen durchführt. Mittelfristig soll hier ein Baukastensystem entstehen, das die Datenanalyse, Bereinigung und Konvertierung, ggf. auch Dublettenerkennung, unabhängig vom Format für beliebige bibliographische Datenbestände in den Bibliotheken ermöglicht.

Die Software

MABLe+ ist ein Programm für die Fehleranalyse von Dateien im MAB-Format. Das Programm wurde bereits erfolgreich bei der Migration von Bibliothekskatalogen in den B3Kat eingesetzt, um vorab die Qualität der Metadaten einschätzen und Probleme beseitigen zu können. MABLe+ läuft auf der Kommandozeile und wurde in Java geschrieben. Das Ergebnis einer MABLe+ Analyse sind vier Textdateien, die verschiedene Statistiken über die gefundenen Daten und Fehler enthalten.

Die Software wird in Kürze unter GPL 3.0 veröffentlicht.

Informationen zu MABLE+

Jens Schwidder
Verbundzentrale des Koope-
rativen Bibliotheksverbundes
Berlin-Brandenburg
Tel.: +49 (0)30-84185-308
E-Mail: schwidder@zib.de

Stefan Brecheisen
Verbundzentrale des
Bibliotheksverbundes Bayern
Tel.: +49 (0)89-28638-2134
E-Mail: stefan.brecheisen@
bsb-muenchen.de

Weiterentwicklung

Für die Weiterentwicklung sind folgende Features ge-
plant:

- ◆ Definition anwenderspezifischer Regeln, um die Analyse den Anforderungen einer Institution an passen zu können,
- ◆ Automatische Korrektur von gefundenen einfachen Fehlern, nach Regeln, die vom Nutzer konfiguriert werden können,
- ◆ Einbeziehung externer Datenquellen, z.B. Linked Open Data (LOD), in die Analyse und Fehlerkorrektur,
- ◆ Unterstützung von ALEPH-Sequential und
- ◆ Unterstützung von MARC 21.

Technologien

MABLE+ ist in Java geschrieben und ist für Windows, Linux und Mac OS-X Systemen optimiert. Es verwendet verschiedene Open-Source-Software-Bibliotheken: Spring Framework, Apache Commons, DDB-Charsets für MAB2 Decoding/Encoding. Bei der Entwicklung werden Maven 3 und JIRA eingesetzt.

Webseite

<http://mable.kobv.de>

```
LDR 00432nM2.01200024-----h
001 134567
010 891011
030 L zz5d||rz
051 s|||||
100a Buss, Harald
106 124255639
330 Feuchteschäden - Umfassungs
410 Kissing [u.a.]
412 WEKA-Verl., Verl. Für Baufa
433 333 S.
451 Aus der Fachbuchreihe Bausch
455 5
461 Betontechnik
463 578329
540a 3-7013-0466-1
```

Beispiel eines fehlerhaften bibliographischen Datensatzes