Metadata Editor

Petr Kovář and Michal Růžička

Masaryk University, Institute of Computer Science, Brno, Czech Republic  
<kovar@ics.muni.cz>

Masaryk University, Faculty of Informatics, Brno, Czech Republic  
<mruzicka@mail.muni.cz>

October 13th, 2010
Metadata Editor

- A tool for digital library metadata / full texts viewing, management, and correcting is needed.

- The YADDA system is not going to integrate a metadata editor.

- The Metadata Editor (tested on the DML-CZ project) is available for the EuDML project.
  - A client–server web application.
  - Designed to
    - manage,
    - edit,
    - and validate each article’s metadata and full texts prior to their integration into the digital library.
The Metadata Editor is intended to be a stand-alone application used by EuDML participants.

- The Metadata Editor is not a content management system (CMS) intended for article preparation, management of reviews of articles, etc.
- The Metadata Editor is designed to be a management and correcting tool for the preparation of full texts and metadata prior to their import to a public digital library.

- Data prepared using the Metadata Editor are then transmitted to the EuDML Core using a designated interface.
Metadata Editor Workflow

- Preparation of input data for the Metadata Editor.
- Load of input data into the Metadata Editor.
- Build of articles from discrete pages (structural metadata).
- Metadata editing (descriptive metadata).
- Bibliographical references (creation, harvesting, linking).
- Automated metadata verification.
- Compilation of final PDFs.
- Export to a publication system.
Input Data

- Digitized old printed documents.
  - Scanned pages + OCR layer.
- Retro-born-digital.
  - Documents were made unaware of the digital library.
  - Conversion of original digital document is necessary.
- Born-digital.
  - Inserted on-line by publishers.
  - Final PDFs + metadata.
Supported Publication Types

- Serials (journal / volume / issue / article).
- Proceedings (series / volume / article).
- Monographs (collection / monograph / chapter).
- Celebrities (celebrity / work type / work).
Interfaces

- Web application:
  - structure creation
  - page handling
  - upload/download of data
  - metadata editing
  - authority database management

- Directory structure:
  - import/export of data
  - compilation of PDFs and other automated operations
  - publishing
Creation of Articles

1. Automated creation of initial data structure.
   - The first and last pages of articles are automatically found and used.

   - Visual article editor.

3. Creation of descriptive metadata.
   - Pre-filled by automated process from citation databases.
<table>
<thead>
<tr>
<th>Article</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>(#1)</td>
<td>Über zwei neue ebene Konfigurationen $(12_4, 16_3)$</td>
<td>193-218</td>
</tr>
<tr>
<td>(#2)</td>
<td>The theory of characters of finite commutative semigroups (30-58)</td>
<td>219-247</td>
</tr>
<tr>
<td>(#3)</td>
<td>System of congruence relations on lattices (59-93)</td>
<td>248-282</td>
</tr>
<tr>
<td>(#4)</td>
<td>Sur les espaces à connexion affine partiellement projectifs (94-101)</td>
<td>283-290</td>
</tr>
<tr>
<td>(#5)</td>
<td>Characters of commutative semigroups as class functions (102-103)</td>
<td>291-292</td>
</tr>
</tbody>
</table>
Page Manipulations

• Page thumbnails.

• Two types of page blocks.
  • Green blocks – pages assigned to an article.
  • Red blocks – pages not assigned to any article.

• Verification of page ordering.

• Reshuffling pages between articles and issues.

• Creation/deletion of articles.

• Page cloning.

• Replacement of page images.

• Page number editing.

• Named sections and subsections.
(#10) --- Název článku nebyl zadán --- 149-(154d)

Save contents
Creation of Descriptive Metadata

- Pre-filled by automated process from citation databases.

- Editing screen has two parts.
  - Left part consist of the editing form.
  - Right part contains preview of the first page.
    - Switching between pages is possible.
O JISTÉ VLASTNOSTI SOUSTAV NEZÁVISLÝCH PRVKŮ V ABELOVSKÉ GRUPĚ

MILAN SEKANINA, Brno
(Doložilo dne 18. července 1969)

V Sílance se dokazuje, že každá neprázdná množina nezávislých prvků z abelovské grupy je jejím faktorem ve smyslu Hajošově.

Nechť \( G \) je abelovská grupa. Neprázdnou podmnožinu \( M \subset G \) nasyťme nezávisle, tedy pro každou neprázdnou konečnou podmnožinu \( N = \{a_1, \ldots, a_n\} \) množiny \( M \), že z rovnice \( v_1a_1 + \cdots + v_na_n = 0 \) (je nulový prvek grupy \( G \)), kde \( v_i \) jsou celá čísla, plyne \( v_i = 0 \) pro \( i = 1, \ldots, n \) (viz [1], str. 123).

Nechť \( M, N \) jsou dvě neprázdné podmnožiny z \( G \). Potom \( M + N \) značí množinu všech těch prvků z \( G \), které se dají psát jako součet prvků z \( M \) a prvků z \( N \). Dále se každý prvek z \( G \) napsat nanejvším jedním způsobem jako \( m + n \), \( m \in M \), \( n \in N \). Píšeme \( M \perp N \). Je-li \( G = M + N \) a \( M \perp N \), píšeme tedy \( M \perp N \) a říkáme, že \( M \) a \( N \) tvoří faktorizaci grupy \( G \) ve smyslu Hajošově (viz též [2]) a \( M \) a \( N \) nasyťme faktory grupy \( G \).

Dokážeme větu:

Věta. Nezávislá množina \( M \subset G \) je faktorem ve smyslu Hajošovi.

Důkaz. I. Nechť \( M \) je konečná množina, tedy \( M = \{a_1, \ldots, a_n\} \). Ukážeme, že

\[ B = \{a_1, \ldots, a_n\} \cup \{k_1a_1 + k_2a_2 + \cdots + k_na_n \mid k_1, k_2, \ldots, k_n \text{ jsou celá čísla} \}, \]

kde \( B \) je nejmenší podgrupa z \( G \) obsahující množinu \( M \), tedy

\[ x \in B \Rightarrow x = h_1a_1 + h_2a_2 + \cdots + h_na_n, \]

kde \( h_1, h_2, \ldots, h_n \) jsou celá čísla (píšeme se též \( B = \{M\} \)). Nechť tedy \( x = h_1a_1 + h_2a_2 + \cdots + h_na_n \) a

\[ h_1 + 2h_2 + \cdots + nh_n = gn + s, \]

kde \( 0 < s \leq n \).

a) Nechť \( s \neq 1 \). Potom \( x = a_1 + n(a_1 - a_1) \), kde \( k_i = h_i \) pro \( 1 \leq i \leq s \), \( h_s = h_1 - 1 \).

338
Authority Database

- Personal metadata.
- Name forms.
  - First name, last name, displayed name form, transliterated name form, attributes.
- Solution for several problems:
  - one person has several name forms
  - more persons have the very same name
  - searching for one form returns all articles created by one person
<table>
<thead>
<tr>
<th>Name</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frunză</td>
<td>Frunză, Ștefan (1)</td>
</tr>
<tr>
<td></td>
<td>Frunză, a, Ș (0)</td>
</tr>
<tr>
<td></td>
<td>Frunză, a, Stefan (0)</td>
</tr>
<tr>
<td></td>
<td>Frunză, a, \c St. (0)</td>
</tr>
<tr>
<td></td>
<td>Frunză, a, \lhook Stefan (0)</td>
</tr>
<tr>
<td>Fryba</td>
<td>Fryba, Ladislav (2)</td>
</tr>
<tr>
<td></td>
<td>Fryba, Ladislav (0)</td>
</tr>
<tr>
<td></td>
<td>Fryba, L. (0)</td>
</tr>
<tr>
<td>Fuchs</td>
<td>Fuchs, Alexander (2)</td>
</tr>
<tr>
<td></td>
<td>Fuchs, Eduard (3)</td>
</tr>
<tr>
<td></td>
<td>Fuchs, E. (0)</td>
</tr>
<tr>
<td></td>
<td>Fuchs, Jaromír (1)</td>
</tr>
<tr>
<td></td>
<td>Fuchs, László (1)</td>
</tr>
<tr>
<td></td>
<td>Fuchs, L. (0)</td>
</tr>
<tr>
<td></td>
<td>Fuchs, Ladislaus (0)</td>
</tr>
<tr>
<td></td>
<td>Fuchs, Laszlo (0)</td>
</tr>
<tr>
<td></td>
<td>Fuks, L. (0)</td>
</tr>
<tr>
<td>FuchsL2</td>
<td>Fuchs, L. (1)</td>
</tr>
<tr>
<td></td>
<td>Fuchs, Lazarus (0)</td>
</tr>
<tr>
<td>FuchsL3</td>
<td>Fuchsová, Libuše (1)</td>
</tr>
<tr>
<td></td>
<td>Fuchsová, Libuše (0)</td>
</tr>
<tr>
<td>FuchsM</td>
<td>Fuchs, Martin (2)</td>
</tr>
<tr>
<td></td>
<td>Fuchs, Martin (0)</td>
</tr>
<tr>
<td></td>
<td>Fuchs, M. (0)</td>
</tr>
<tr>
<td>Fuchk</td>
<td>Fučik, Josef (1)</td>
</tr>
<tr>
<td></td>
<td>Fučik, Josef (0)</td>
</tr>
<tr>
<td>FuchkS</td>
<td>Fučik, Svatoopluk (33)</td>
</tr>
<tr>
<td></td>
<td>Fučik, Svatoopluk (0)</td>
</tr>
<tr>
<td></td>
<td>Fučik, S. (0)</td>
</tr>
<tr>
<td></td>
<td>Fučik, Svatoopluk (0)</td>
</tr>
<tr>
<td></td>
<td>Fučik, S. (0)</td>
</tr>
<tr>
<td>Fuhni</td>
<td>Fuhrig, Josef (1)</td>
</tr>
<tr>
<td>Fukal</td>
<td>Fuka, Jaroslav (9)</td>
</tr>
<tr>
<td></td>
<td>Fuka, J. (1)</td>
</tr>
<tr>
<td></td>
<td>Fuka, Ja. (0)</td>
</tr>
<tr>
<td>FulksW</td>
<td>Fulks, Watson (1)</td>
</tr>
<tr>
<td></td>
<td>Fulks, W. (0)</td>
</tr>
<tr>
<td></td>
<td>Fulks, W. B. (0)</td>
</tr>
</tbody>
</table>
PDF Compilation

• The final PDF consists of:
  • cover page (generated automatically using \TeX)
  • the article itself
  • OCR layer (scanned pages only)
  • digital signature
Dynamic Generation of Editing Forms (SForms Project)

• One of the most important functions of the Metadata Editor consists in facilitating interactive modification of metadata.

• The operators are allowed to browse the contents of the repository and make necessary adjustments through the web-based interface of the relevant forms.

• The metadata language is formally defined by an XML Schema \(\Rightarrow\) it is possible to generate the forms dynamically based on the XML Schema definition.
Dynamic Generation of Editing Forms (SForms Project) (cont.)

- The mechanism consists of server-side and client-side scripting.
  - The XML Schema is enriched with hints for visualising HTML form and mapping to core-elements.
  - The XML Schema is processed on the server by a Perl script.
  - The script generates the JavaScript code that is included in the web page and which is subsequently sent to the client.
  - This JavaScript code runs in the web browser of the end user and generates a form that matches the language defined by the source XML Schema.

- Not all features of the XML Schema are supported, but the mechanism is powerful enough to satisfy the requirements.

- A generalized version of the forms generator is available as a stand-alone open-source project.
On-line Submissions and Validation

• The viability of a digital library rests with new acquisitions emerging mainly in the form of born-digital publications.

• The born-digital inputs to the Metadata Editor come from different sources, primarily from editors of various journals. ⇒ There is need for data validation.

• \( \text{TEX} \) codes have to be correct.
  • The digital library could use the metadata to generate \( \text{TEX} \) document.
  • Conversion to MathML.

• The editors themselves need feedback ⇒ on-line application.
  • Make it possible to validate metadata during preparation of a new journal issue.
  • Make it possible to submit final version of data to the digital library.
On-line Submissions and Validation (cont.)

<meta.xml:6: element keyword: Schemas validity error : Element 'keyword': This element is not expected. Expected is one of ( author, language ).

meta.xml fails to validate

#####

##### CHYBA: Clankova metadata "meta.xml" v adresari "/1/" nejsou validni.

#####
On-line Submissions and Validation (cont.)

meta.xml:6: element keyword: Schemas validity error : Element 'keyword': This element is not expected. Expected is one of ( author, language ).
meta.xml fails to validate

Manage

Issue is valid
Automated Metadata Verification

• On upload.
  • When publisher upload a new issue.

• On demand.
  • Run one of the tests by administrator (can be done for a specific journal or issue).

• Automated.
  • Planned verifications for the whole database.
  • Results available from the Metadata Editor.

• Some of the tests (modular plug-in scripts).
  • XML validation, missing metadata, article vs. OCR language, syntax of TeX expressions.
Overall Architecture

- Data and metadata are stored in directory structure.
- Web interface index metadata in MySQL.
  - Apache.
  - Ruby (ramaze and og frameworks).
  - External scripts (Perl, Bash, …).
- Automated actions (OCR, verification, compilation of PDFs).
- Publishing.
Regular Files vs. Relational Database

- Data and metadata are stored in regular files.
- Relational database is used to speed up some operations such as search.
  - The database contains only a copy of the data.
- The database can be dropped and recreated again from the files in case of problems.
- It is easier to backup/transfer regular files.
- It is easy to view/edit metadata even with basic tools such as text editor.
Documentation

• Metadata Editor contains integrated on-line help.

• The help system uses README files from the Metadata Editor directory structure as source files.
  
  • Markdown markup language (wiki-like plaintext syntax) is used.
  
  • Each of the README files describe content of the directory which it is placed in.
  
  • Help is available from the Metadata Editor as well as from the Metadata Editor sources / installation directory.
    
    • Documentation is easy to manage: information shown in the integrated on-line help system is the very same as the README files.

• Annotated example configuration, metadata, etc. files are part of the Metadata Editor sources. They are integrated with the help system.
Internationalization

• To make the Metadata Editor useful for international projects, the Editor application was translated and localized.

• Adapting the user interface of an existing application to new languages involves changing the output in a way that will please the current user.

• The Metadata Editor contains integrated translation tool.
  • English and Czech versions are now available.
  • Additional translations will be add as needed.
## Translation to Czech completed from 83% (5/6)

<table>
<thead>
<tr>
<th>Příznak</th>
<th>Czech</th>
<th>Výchozí hodnota</th>
<th>Klíč</th>
<th>Komentář překladatele</th>
</tr>
</thead>
<tbody>
<tr>
<td>ok</td>
<td>přihlašovací jméno</td>
<td>login</td>
<td>user.change.login</td>
<td></td>
</tr>
<tr>
<td>Uložit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ok</td>
<td>nové heslo</td>
<td>password</td>
<td>user.change.password.new</td>
<td></td>
</tr>
<tr>
<td>Uložit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nepřeložený</td>
<td></td>
<td>old password</td>
<td>user.change.password.old_default</td>
<td></td>
</tr>
<tr>
<td>Uložit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ok</td>
<td>znovu nové heslo</td>
<td>repeat password</td>
<td>user.change.password.repeat</td>
<td></td>
</tr>
<tr>
<td>Uložit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ok</td>
<td>Změnit heslo</td>
<td>Change password</td>
<td>user.change.submit</td>
<td></td>
</tr>
<tr>
<td>Uložit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ok</td>
<td>Aktualizovat profil</td>
<td>Update profile</td>
<td>user.change.update_profile</td>
<td></td>
</tr>
<tr>
<td>Uložit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Output Interface

- Internal data format of the Metadata Editor are regular files in hierarchical directory structure.
  - PDF full texts.
  - XML metadata.

- It is easy to use XSLT for metadata transformations.
  - Full contents of the DML-CZ digital library was successfully transformed to the NLM format.

- It is possible to add an interface providing data in any desired format.
  - REPOX.
Conclusions

• The Editor is now in use in a variety of environments. These include:
  • the DML-CZ project
  • the Faculty of Arts of Masaryk University
  • the Kramerius project of the Moravian Library

• The Editor is used by the EuDML project participants:
  • Czech Republic – Masaryk University
  • Bulgaria – Institute of Mathematics and Informatics
  • Greece – Department of Informatics of the Ionian University
Conclusions (cont.)

- The Metadata Editor is a live, continuously developing project. New features are added as needed.

- The on-line input and validation service was worked in to provide users with a comfortable and safe interface for data inclusion.

- The user interface is dynamically generated based on the formal definition of the metadata.

- The Metadata Editor is used in several projects including the EuDML project participants.
Metadata Editor Sandbox

<http://test.editor.dml.cz/>

login: admin
password: admin
Questions?
Czech Digital Mathematics Library [online].
[cit. 2010-10-09].

Digitization Metadata Editor [online].
[cit. 2010-10-09].
Available from WWW: <http://dme.sourceforge.net/>

Bartošek, M., Kovář, P., Šárfy, M.:
DML-CZ Metadata Editor : Content Creation System for Digital Libraries.

Šárfy, M.:
Metadatový editor pro digitální knihovny.

Jiří Rákosník:
DML-CZ: Asks and Bids.

EuDML: The European Digital Mathematics Library [online].
This page was last modified on 20 January 2010, at 08:09. [cit. 2010-10-09].

SchemaForms [online].
[cit. 2010-10-09].
Available from WWW: <http://sforms.sourceforge.net/>