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

EuDML
The EUROPEAN DIGITAL MATHEMATICS LIBRARY

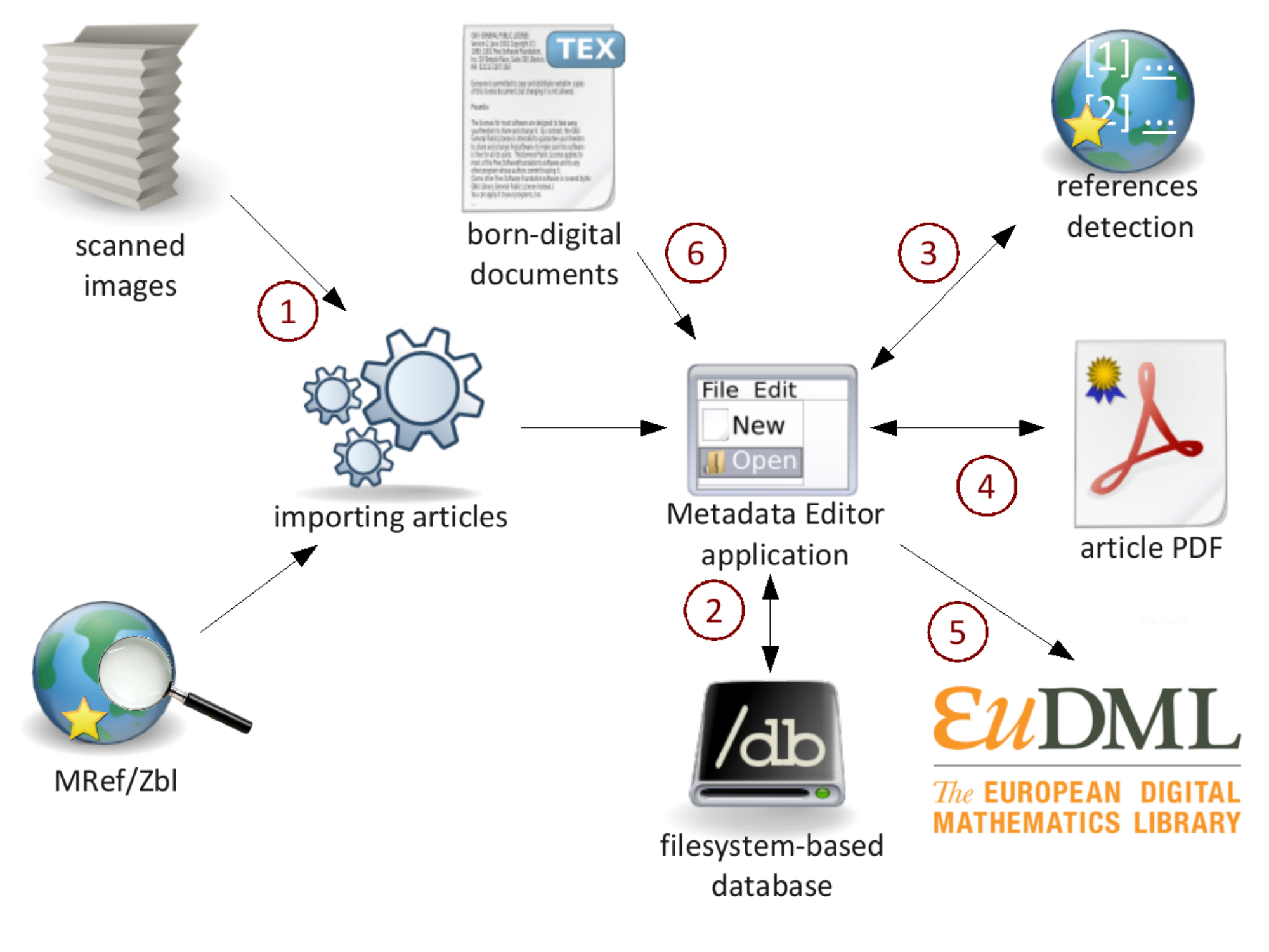
Metadata Editor

- 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.
  - Open-source software (<http://dme.sourceforge.net/>).

Metadata Editor (cont.)

- 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.


scanned images
born-digital documents
TEX
references detection
[1] ...
[2] ...
1
6
3
importing articles
File Edit
New
Open
Metadata Editor application
4
article PDF
MRef/Zbl
2
5
/db
filesystem-based database
EuDML
The EUROPEAN DIGITAL MATHEMATICS LIBRARY

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.
2. Manual checks and corrections of pagination.
   - Visual article editor.
3. Creation of descriptive metadata.
   - Pre-filled by automated process from citation databases.

DML-CZ: Metadata editor (serial)

Delete Articles | Change Ranges | Save Contents

(193a) [2] (193b) [3]

edit ocr scan edit ocr scan

Move Pages | Create Article

(#1) Über zwei neue ebene Konfigurationen $(12_4, 16_3)$ 193-218

Obsah

193 [4] 194 [5] 195 [6] 196 [7] 197 [8]

edit ocr scan edit ocr scan edit ocr scan edit ocr scan edit ocr scan

198 [9] 199 [10] 200 [11] 201 [12] 202 [13]

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.


edit ocr scan


edit ocr scan


edit ocr scan

Remove Pages | Move Pages | Create Article | Delete Article

(#10) --- Název článku nebyl zadán ---

149-(154d)

go Top | Check : All : None : Invert

149 [74]


edit ocr scan

150 [75]


edit ocr scan

151 [76]


edit ocr scan

152 [77]


edit ocr scan

153 [78]


edit ocr scan

154 [79]


edit ocr scan

(154a) [80]


edit ocr scan

(154b) [81]


edit ocr scan

(154c) [82]


edit ocr scan

(154d) [83]


edit ocr scan

Remove Pages | Move Pages | Create Article | Delete Article

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.

Save | Save and Next

338 prev | next | references
339
340
341

<< | < | > | >>

Status: in progress

Title: On a certain property of a set of independent elements of — English

Title: O jisté vlastnosti soustav nezávislých prvků v abelovské gr — Czech

Title:

Author: Sekanina, Milan — SekanM

Author:

Language: Czech

Language:

Keywords:

Summary:

Summary Language: English

Summary Language: Russian

Summary Language:

MSC: 20-30

MSC:

idMR: MR0123606 Mathematical Reviews

idZBL: Zbl 0122.03601 Zentralblatt MATH

Časopis pro pěstování matematiky, roč. 85 (1960), Praha

O JISTÉ VLASTNOSTI SOUSTAV NEZÁVISLÝCH PRVKŮ V ABELOVSKÉ GRUPĚ

Milan Sekanina, Brno

(Došlo dne 18. července 1959)

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

Nechť $\mathfrak{G}$ je abelovská grupa. Neprázdnou podmnožinu M z $\mathfrak{G}$ nazýváme nezávislou, platí-li pro každou neprázdnou konečnou podmnožinu $N = \{a_1, \ldots, a_n\}$ množiny M, že z rovnice $v_1 a_1 + \ldots + v_n a_n = 0$ (0 je nulový prvek grupy $\mathfrak{G}$), 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 $\mathfrak{G}$. Potom $M + N$ značí množinu všech těch prvků z $\mathfrak{G}$, které se dají psát jako součet prvku z M a prvku z N. Dá-li se každý prvek x z $\mathfrak{G}$ psát nanejvýš jedním způsobem jako $m + n$, $m \in M$, $n \in N$, píšeme $M \perp N$. Je-li $\mathfrak{G} = M + N$ a $M \perp N$, píšeme též $M \dot{+} N$ a říkáme, že M a N tvoří faktorisaci grupy $\mathfrak{G}$ ve smyslu Hajósově (viz též [2]) a M a N nazýváme faktory grupy $\mathfrak{G}$.

Dokážeme větu:

Věta. *Nezávislá množina $M \subset \mathfrak{G}$ je faktorem $\mathfrak{G}$ ve smyslu Hajósově.*

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

$$\mathfrak{M} = \{a_1, \ldots, a_n\} \dot{+} \mathbf{E}[k_1 n a_1 + k_2(a_2 - 2a_1) + \ldots + k_n(a_n - n a_1)\,,$$
$$k_1, k_2, \ldots, k_n \text{ probíhají množinu celých čísel}]\,,$$

kde $\mathfrak{M}$ je nejmenší podgrupa z $\mathfrak{G}$ obsahující množinu M, tedy

$$x \in \mathfrak{M} \Leftrightarrow x = h_1 a_1 + h_2 a_2 + \ldots + h_n a_n\,,$$

h_i celé číslo (píše se též $\mathfrak{M} = [M]$). Nechť tedy $x = h_1 a_1 + h_2 a_2 + \ldots + h_n a_n$ a

$$h_1 + 2h_2 + \ldots + n h_n = gn + s\,,$$

kde $0 < s \leqq n$.

a) Nechť $s \neq 1$. Potom $x = a_s + n g a_1 + \sum_{i=2}^{n} k_i(a_i - i a_1)$, kde $k_i = h_i$ pro $i \neq s$, $k_s = h_s - 1$.

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

☐	FrunzS (1)	Frunză, Ştefan (1) Frunz\u a, \c Stefan (0) Frunz\u a, S. (0) Frunz\u a, Stefan (0) Frunz\u a, \c St. (0) Frunz\u a, \lfhook Stefan (0)	Join Delete
☐	FrybaL (2)	Frýba, Ladislav (2) Fr\'yba, Ladislav (0) Fr\'yba, L. (0)	Join Delete
☐	FuchsA (2)	Fuchs, Alexander (2)	Join Delete
☐	FuchsE (3)	Fuchs, Eduard (3) Fuchs, E. (0)	Join Delete
☐	FuchsJ (1)	Fuchs, Jaromír (1)	Join Delete
☐	FuchsL (1)	Fuchs, László (1) Fuchs, L. (0) Fuchs, Ladislaus (0) Fuchs, Laszlo (0) Fuks, L. (0)	Join Delete
☐	FuchsL2 (1)	Fuchs, L. (1) Fuchs, Lazarus (0)	Join Delete
☐	FuchsL3 (1)	Fuchsová, Libuše (1) Fuchsová, Libu\v se (0)	Join Delete
☐	FuchsM (2)	Fuchs, Martin (2) Fuch, Martin (0) Fuchs, M. (0)	Join Delete
☐	FucikJ (1)	Fučík, Josef (1) Fu\v cík, Josef (0)	Join Delete
☐	FucikS (33)	Fučík, Svatopluk (33) Fu\v cík, Svatopluk (0) Fuchik, S. (0) Fu\v cik, Svatopluk (0) Fu\v cík, S. (0) {\cyr Fuchik, S.} (0)	Join Delete
☐	FuhriJ2 (1)	Fuhrig, Josef (1)	Join Delete
☐	FukaJ (10)	Fuka, Jaroslav (9) Fuka, J. (1) Fuka, Ja. (0)	Join Delete
☐	FulksW (1)	Fulks, Watson (1) Fulks, W. (0) Fulks, W. B. (0)	Join Delete

Journals

General Topology and its Relations to Modern Analysis and Algebra (eng)
Applications of Mathematics (cze)
Archivum Mathematicum (eng)
Archivum Mathematicum (retro) (cze)
Časopis pro pěstování matematiky (cze)
Časopis pro pěstování matematiky a fysiky (cze)
pomocne Časopis pro pěstování matematiky a fysiky - JENSTEJN (cze)
pomocne Commentationes mathematicae Universitatis Carolinae - JENSTEJN (cze)
Commentationes Mathematicae Universitatis Carolinae (eng)
Czechoslovak Mathematical Journal (eng)

Query

function

Element

All : None : Invert

- ☑ Title
- ☑ Author
- ☑ MSC
- ☑ idMR
- ☑ idZBL
- ☑ idJFM
- ☑ Note
- ☑ Note: private
- ☑ Error
- ☑ Language

Property

Language: Any

Category

All : None : Invert

- ☑ math
- ☑ news
- ☑ history
- ☑ politics
- ☑ editorial
- ☑ contents
- ☑ other
- ☑ review
- ☑ physics

Relation

- ◉ equal to
- ○ none equal to
- ○ empty
- ○ not empty
- ○ exact

Display

All : None : Invert

- ☑ Title
- ☑ Author
- ☑ MSC
- ☑ idMR
- ☑ idZBL
- ☑ idJFM
- ☑ Note
- ☑ Note: private
- ☑ Error
- ☑ Language

Search | Reset Form

MSC

Title

Append

159 articles matches

Article ID	Type	Title	Author	MSC	idMR	idZBL	idJFM	Note	Note: private	Error	Language
serial/CzechMathJ/29-1979-3/3	math	On the differentiation of convex functions in finite and infinite dimensional spaces (eng)	Zajíček, Luděk	52A05 46G05 26A27	MR536060	0429.46007					eng

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.

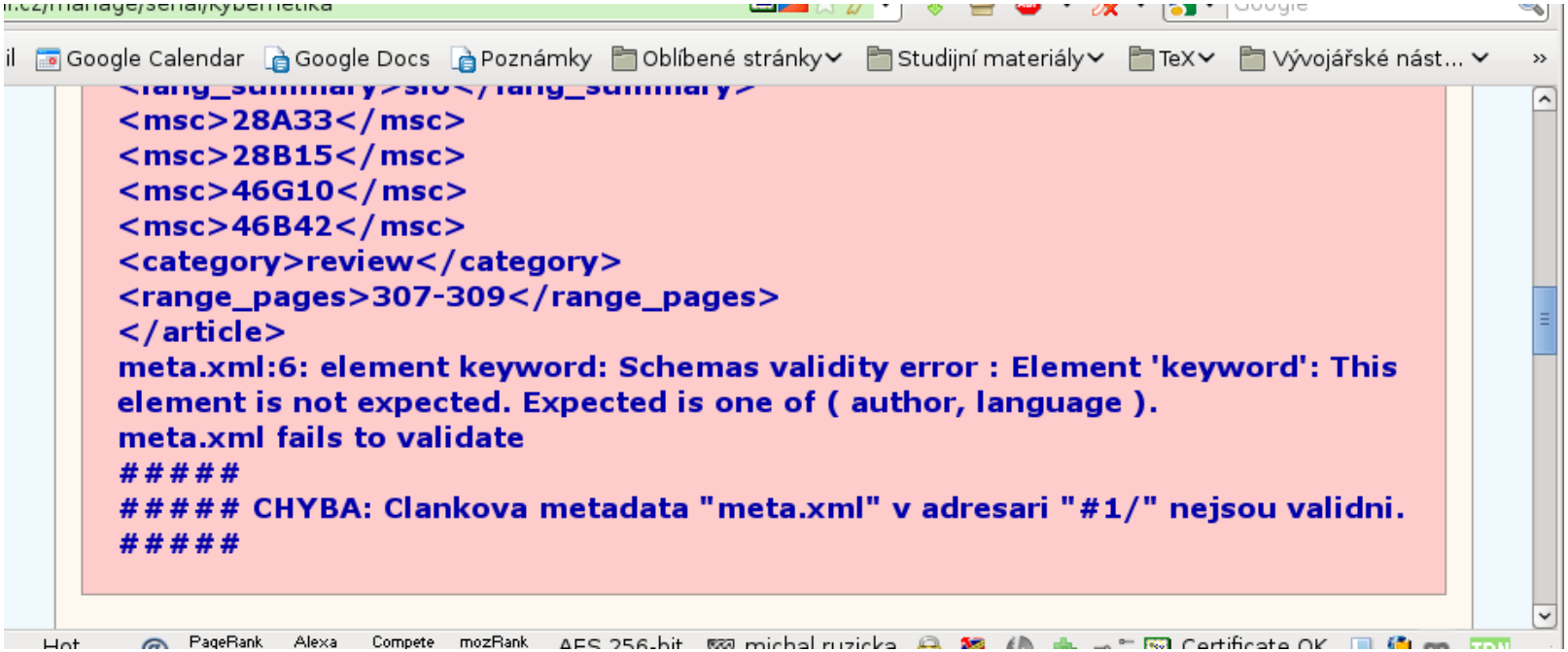
Dynamic Generation of Editing Forms (SForms Project) (cont.)

- 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.

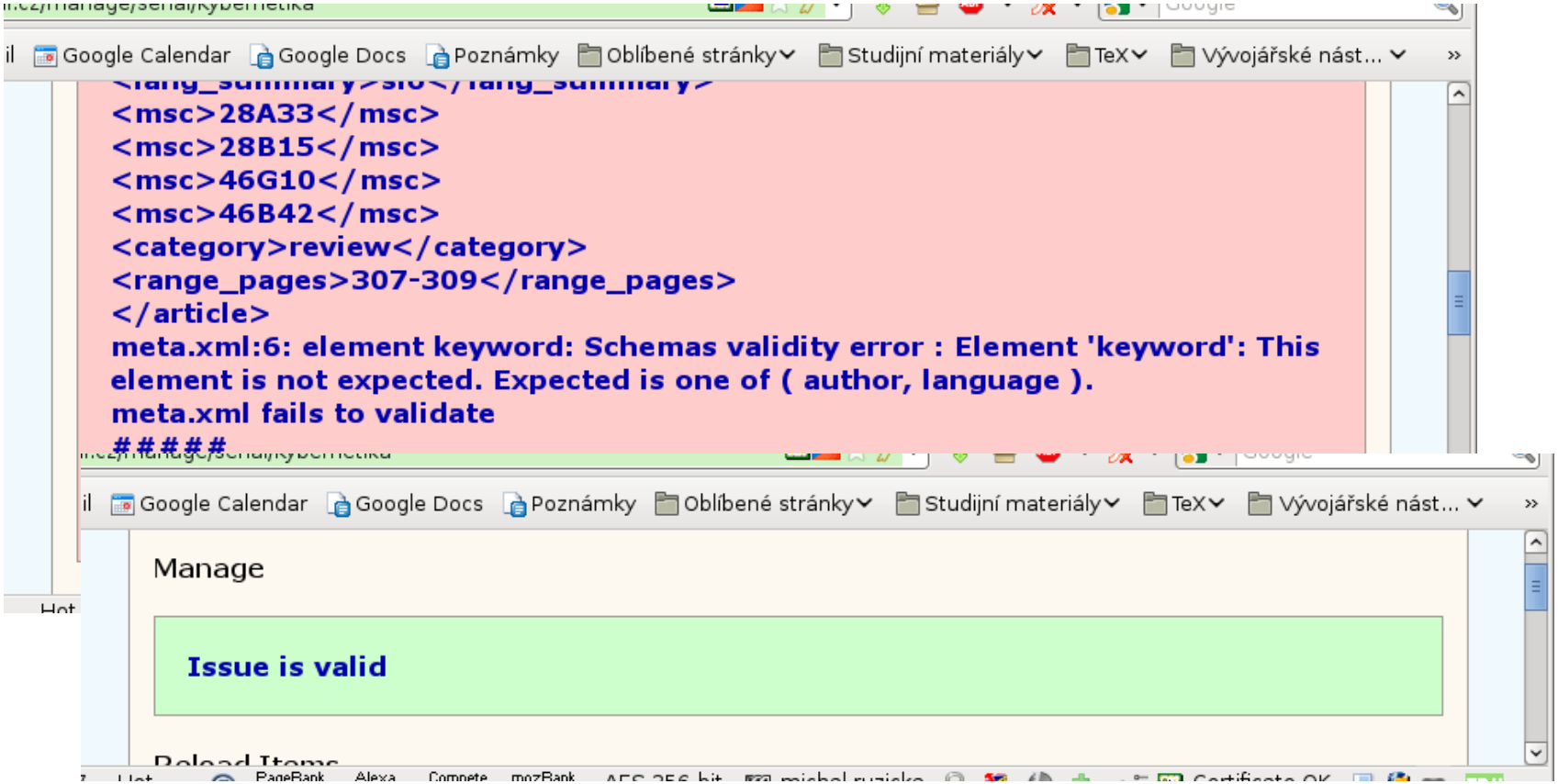
On-line Submissions and Validation

- 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. $\Rightarrow$ There is need for data validation.
- TEX codes have to be correct.
  - The digital library could use the metadata to generate TEX document.
  - Conversion to MathML.
- The editors themselves need feedback $\Rightarrow$ 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.)



On-line Submissions and Validation (cont.)



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.

DML-CZ: Metadatový editor

Translation to Czech completed from 83% (5/6)

Edit file template/user/change/cze.po change file

Příznak	Czech	Výchozí hodnota	Klíč	Komentář překladatele
ok Uložit	přihlašovací jméno	login	user.change.login	
ok Uložit	nové heslo	password	user.change.password.new	
nepřeložený Uložit		old password	user.change.password.old_default	
ok Uložit	znovu nové heslo	repeat password	user.change.password.repeat	
ok Uložit	Změnit heslo	Change password	user.change.submit	
ok Uložit	Aktualizovat profil	Update profile	user.change.update_profile	

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.

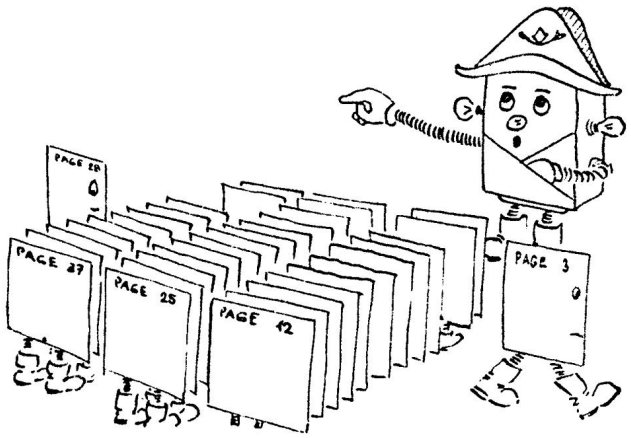
Metadata Editor Sandbox

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

login: admin

password: admin

Questions?




Czech Digital Mathematics Library [online].
[cit. 2010-10-09].
Available from WWW: <http://dml.cz/>.


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.
In: Sojka Petr (editor): *DML 2008 – Towards Digital Mathematics Library*, Birmingham, UK, July 27th, 2008, 139–151.


Šárfy, M.:
Metadatový editor pro digitální knihovny.
In: *Knihovny současnosti 2009*, Seč u Chrudimi, CZ, June 23th, 2009, 140–154. Brno, 2009, pp. 207. ISBN 978-80-86249-54-4
Available from WWW: <http://www.sdruk.cz/sec/2009/sbornik/2009-6-140.pdf>


Jiří Rákosník:
DML-CZ: Asks and Bids.
Presentation at Preparatory Meeting ‘Towards a European Virtual Library in Mathematics’ (European Science Foundation). Santiago de Compostela, Spain, March 13th–14th, 2009.


EuDML: The European Digital Mathematics Library [online].
This page was last modified on 20 January 2010, at 08:09. [cit. 2010-10-09].
Available from WWW: <http://www.eudml.eu/>.


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