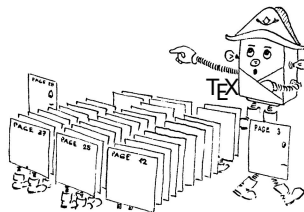


# TeXové technologie pro digitální matematickou knihovnu

Petr Sojka, Michal Růžička

<sojka@fi.muni.cz>, <mruzicka@mail.muni.cz> (FI MU, Brno)

přednáška pro CSTUG, 21. listopadu 2009



## Vize WDML/EuDML

Na počátku byla představa veškerého matematického *recenzovaného* vědění (100 000 000 stran) na jednom místě a v digitální podobě.

Začíná se naplňovat, ale pomalu: tříletý pilotní EU projekt EuDML (7. rámcový program) února 2010 (MU a MU AV).

Základem dosavadní repozitáře jako DML-CZ nebo NUMDAM (budování DL zespodu).

Ukázka DML-CZ (kdo, co, jak brouzdat, jak podobné, jak hledat).

Výzva k připomínkám a algoritmům podobnosti.

# Data

Proof. Let  $\hat{K}$  be a cube,  $\hat{K} \subset \hat{G}$ ; put  $K = \varphi^{-1}(\hat{K})$ . According to theorem 50 we have  $K \in \mathfrak{M}$  and it follows from theorem 24 that

$$P(K, \nu) = \int_K f(x) \, dx. \quad (89)$$

The functional determinant  $T$  of the mapping  $\varphi = \varphi^{-1}$  fulfils the relation  $T(\varphi(x)) \cdot \det M(x) = 1$ , so that

$$\int_K f(x) \, dx = \int_{\hat{K}} f(\varphi(y)) \cdot |T(y)| \, dy = \int_{\hat{K}} \hat{f}(y) \, dy. \quad (90)$$

From theorem 50 (and relation (86)) we see that  $P(K, \nu) = P(\hat{K}, \hat{\nu})$ ; relations (89), (90) show therefore that  $P(\hat{K}, \hat{\nu}) = \int_{\hat{K}} \hat{f}(y) \, dy$ , which completes the proof.

Remark. The reader may compare this paper with [6].

## REFERENCES

- [1] F. Jarník: *Diferenciální počet*, Praha 1953.
- [2] F. Jarník: *Integrální počet II*, Praha 1955.
- [3] J. Mařík: Vrcholy jednotkové koule v prostoru funkcí na daném poloopoběžném prostoru, *Časopis pro příst. mat.*, 79 (1954), 3—40.
- [4] Ян Маржик (Jan Mařík): Представление функционала в виде интеграла, *Чехословацкий мат. журнал*, 5 (80), 1956, 467—487.
- [5] J. Mařík: Plošný integrál, *Časopis pro příst. mat.*, 81 (1956), 79—82.
- [6] Ян Маржик (Jan Mařík): Заметки к теории поверхностного интеграла, *Чехословацкий мат. журнал*, 5 (81), 1956, 387—400.
- [7] S. Saks: *Theory of the integral*, New York.

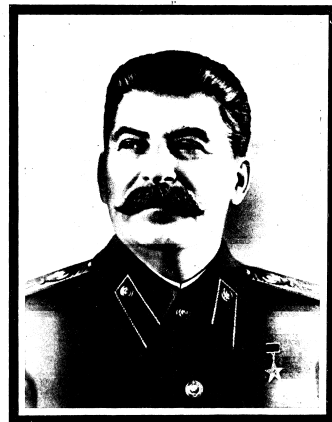
## Резюме

### ПОВЕРХНОСТНЫЙ ИНТЕГРАЛ

ЯН МАРЖИК (Jan Mařík), Прага.

(Поступило в редакцию 10/X 1955 г.)

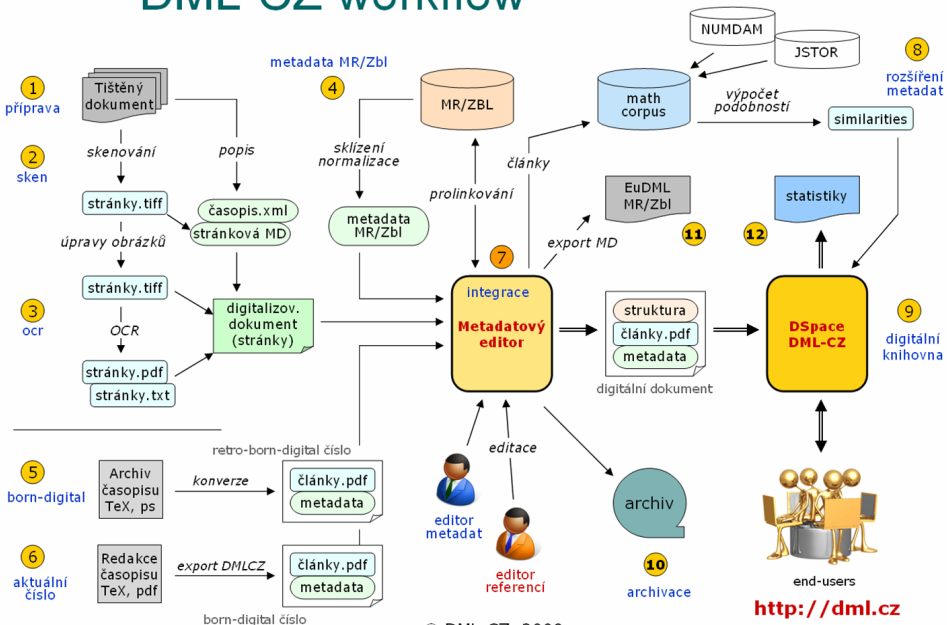
Пусть  $m$  — натуральное число; пусть  $E_m$  —  $m$ -мерное евклидово пространство. Для всякого ограниченного измеримого множества  $A \subset E_m$  положим  $\|A\| = \sup \int_A \sum_{i=1}^m \frac{\partial v_i(x)}{\partial x_i} \, dx$ , где  $v_1, \dots, v_m$  — многочлены такие, что  $\sum_{i=1}^m v_i^2(x) \leq 1$  для всех  $x \in A$ . Пусть  $\mathfrak{M}$  — система всех ограниченных измеримых множеств  $A$ , для которых  $\|A\| < \infty$ . Теорема 18 тогда утверждает: Пусть  $A \in \mathfrak{M}$ ; пусть  $D$  — граница множества  $A$ . Тогда на системе  $\mathfrak{M}$  всех борелевских подмножеств множества  $D$  существует мера  $\rho$  и на



ИОСИФ ВИССАРИОНОВИЧ СТАЛИН

1879—1953

# DML-CZ workflow



© DML-CZ, 2009

# Take care!



# Heterogenita dat a různost formátů

období retro-digital: sken, geometrické transformace (BookRestorer),  
OCR (FineReader, InftyReader), dvouvrstvé PDF

období retro-born-digital: neúplná .tex či .dvi data, špatné formáty,  
bitmapová písma nízkého rozlišení

data z období born-digital: sazba  $\TeX$ em s exportem do metadat digitální  
knihovny

svět autorský:  $\LaTeX$ ,  $\TeX$ ová notace matematiky

svět aplikací: XML, MathML

## Sazba článků a předsádek

- Xe $\LaTeX$ , Charis SIL (množství abeced a znaků ve jménech autorů, cyrilice, ...)
- `\usepackage{pdfpages}` nebo `pdftk` (zachování anotací).
- $\TeX$ ový zdroják generován z XML metadat (XSLT a perl), po validaci metadat plně automatické přegenerování `meta.xml`  $\rightarrow$  `item.xml`  $\rightarrow$  `item.tex`  $\rightarrow$  `item.pdf`  $\rightarrow$  ...

# meta.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<article>
  <number>1</number>
  <status>completed</status>
  <title lang="fre">Sur quelques applications des dispersions centrales da
  <title lang="eng">On some applications of cetral dispersions in the theo
  <author id="Boruv0" order="1">Borůvka, Otakar</author>
  <language>fre</language>
  <msc>34C10</msc>
  <idMR>MR0197823</idMR>
  <idZBL>Zbl 0151.10804</idZBL>
  <idUlrych>19650001</idUlrych>
  <category>math</category>
  <range>7-26</range>
  <range_pages>1-20</range_pages>
  <access>true</access>
</article>

```



## item.tex

```
\newlength{\vsx} \vsx=148mm
\newlength{\vsy} \vsy=205mm
\newcommand\toptitle{Archivum Mathematicum}
\newcommand\maintitle{Sur quelques applications des dispersions centrales}
\newcommand\mainauthors{Otakar Borůvka}
\newcommand\PURL{http://dml.cz/dmlcz/104576}
\documentclass{dmlcz}
\begin{document}
\copyrightholders{$\copyright$ Masaryk University, 1965}
\biptoks{\textit{Archivum Mathematicum},
Vol. 1 (1965), No. 1, 1--20}

\dmltitlepage
\dmlpage{../page/0007}{121mm}{193mm}
\dmlpage{../page/0008}{118mm}{189mm}
\dmlpage{../page/0009}{118mm}{186mm}
...
\end{document}
```

## Další ověřené technologie

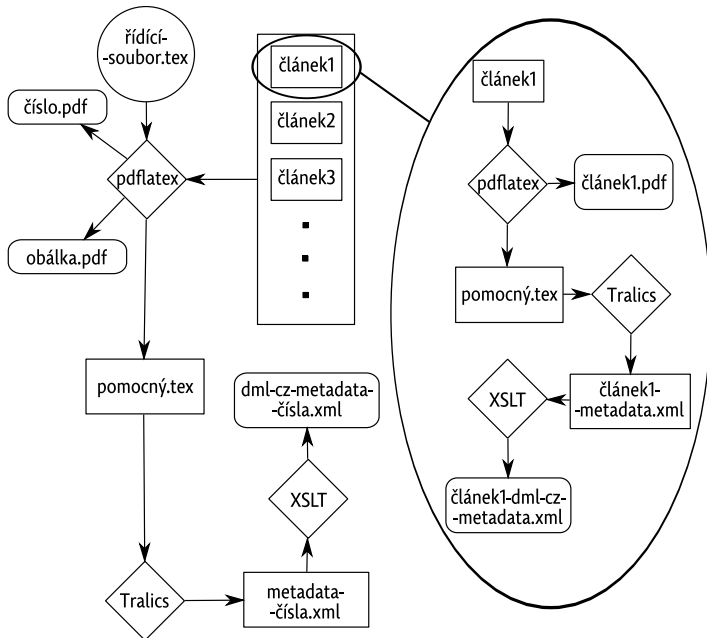
- transformace obrazu (BP Pulkrábek).
- rozpoznávání matematických textů: OCR (DP Panák, Mudrák, BP Vystrčil).
- podepisování PDF: pdfsign (BP Peter Bočák).
- vzdálená editace metadat: webová aplikace metadatový editor (ÚVT MU).
- optimalizace PDF: pdfopt (z ghostscript).
- rekompresi JBIG2: aplikace využívající jbig2enc (BP Radim Hatlapatka).
- počítání podobných článků (výzkum s Radimem Řehůrkem).
- fixfont (BP “wanted!”).
- born-digital redakční systém (BP Michal Růžička).
- ...

# Born-digital časopisy DML-CZ

- Hlavní myšlenka: Získávat born-digital data pro DML-CZ jako vedlejší efekt práce redakcí při vydávání nových čísel časopisů.
- Snaha automatizovat co nejvíce činností...
  - Komplexní redakční systém Archiva Mathematica.
- ...ale nekomplikovat zaběhnutý pracovní cyklus redakce.
  - Minimalistický systém pro získávání DML-CZ metadat.

# Redakční systém časopisu Archivum Mathematicum

- Inspirován francouzským systémem CEDRAM.
- Pro generování XML metadat použít Tralics.
- Články ve formátu  $\text{\LaTeX}$  s AMS styly, Bib $\text{\TeX}$ .
- Automatické vytvoření celého čísla ze zadaných článků, vytvoření elektronické a tiskové verze, další pomocné operace.
- Automatizované předání dat projektu DML-CZ.



# Co je Tralics

- Konvertor  $\text{\LaTeX}$ u do XML.
- Multiplatformní (GNU/Linux, Apple MacOS X, Microsoft Windows).
- Free software, CeCILL licence ( $\approx$  GNU GPL).
- Domovská stránka: `<http://www-sop.inria.fr/apics/tralics/>`

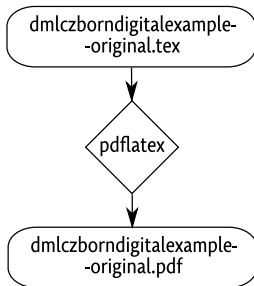
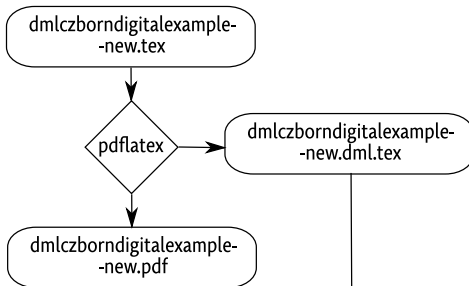
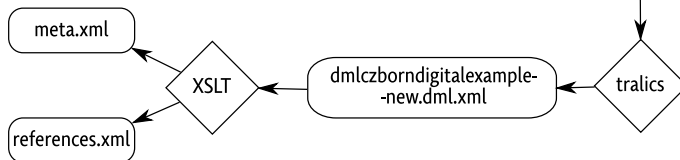
# Jak Tralics pracuje

- Sám Tralics rozumí zdrojovému textu  $\text{T}_{\text{E}}\text{X}$ u a dokáže jej přeložit.
  - Pracuje přímo s  $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ ovými zdrojovými texty.
  - Velmi kvalitní zpracování vstupních zdrojových textů.
  - Přirozená konfigurace.
- Dokáže zpracovávat bibliografické databáze Bib $\text{T}_{\text{E}}\text{X}$ u.

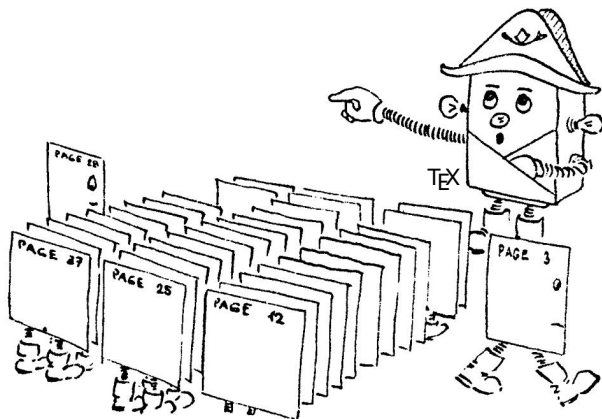
# Konfigurace Tralicsu

- Tralics přidává nové příkazy  $\TeX$ u, kterými je možné řídit transformaci obsahu dokumentu do XML.
  - V konfiguračních souborech Tralicsu definujeme chování značkování použitého v dokumentu při překladu do XML.
- Přirozená struktura konfigurace –  $\LaTeX$ ová třída dokumentu / balíček (.cls/.sty) má obdobu v konfiguračním souboru pro Tralics (.clt/.plt).
  - Připojení další konfigurace specifické pro konkrétní dokument je také snadné.
- Součástí distribuce jsou předpřipravené konfigurační soubory pro základní třídy dokumentů /  $\LaTeX$ ové balíčky.



make originalmake newmake metadata

# Dotazy?



# Seznam použité literatury a odkazy



## APICS TEAM.

*Tralics: a LaTeX to XML translator* [online].

Last modified \$Date: 2009/10/15 09:25:34 \$ [cit. 2009-11-14].

Dostupný z WWW: <<http://www-sop.inria.fr/apics/tralics/>>.



## BOUCHE, THIERRY.

*CEDRICS: When CEDRAM Meets Tralics.*

In: Sojka Petr (editor): DML 2008 – Towards Digital Mathematics Library, Birmingham, UK, July 27<sup>th</sup>, 2008, 153–165.



## DML-CZ TÝM.

*Materiály projektu DML-CZ* [online, cit. 2009-11-14].

<<http://project.dml.cz>>.