The Czech Digital Mathematics Library project

Project Summary

Funding body *Academy of Sciences of the Czech Republic*
Funding program *"Information Society" (National Research Programme, 2005–2009)*

Introduction

A necessary pre-condition for research and development in every field of science is efficient access to information in scholarly literature. Together with the rapidly increasing extent of this kind of information, a strong demand exists for an efficient means of storing, organizing, searching and accessing information in a digital environment. This is particularly true in mathematics which serves as a core and important source of knowledge for a wide range of research and development in all sciences.

Objectives

The aim of this project is to investigate, develop and apply techniques, methods and tools that would allow the creation of a suitable infrastructure and conditions for establishing what will become the Czech Digital Mathematics Library (DML-CZ). It will consist of the relevant mathematical literature which has been published throughout history in the Czech lands. Upon completion it will be incorporated into the World Digital Mathematics Library (WDML).

The project will involve launching the digitization process and providing end users with access to the digitized material. It will also involve research into advanced technologies for searching mathematical documents, and for including born-digital materials.

The DML-CZ should primarily contain professional journals of international standing published by Czech institutions, such as *the Czechoslovak Mathematical Journal* and *Applications of Mathematics* published by the Mathematical Institute of the Czech Academy of Sciences, *Kybernetika* published by the Institute of Automation and Information Theory AS CR and others. The DML-CZ will also include conference proceedings published by Czech universities and research institutes, selected monographs, textbooks, dissertation theses and research reports. The creation of the DML-CZ will lay the foundations for the Czech component of the WDML into which it will be integrated.

Proposed Solution

The first (preparatory) phase will involve an investigation of the state-of-the-art knowledge and trends in digitizing mathematical literature, and building a digital environment for mathematics (not to reinvent the wheel). On this basis solutions will be proposed and implemented reflecting Czech specifics and settings and, at the same time, taking into account world standards and trends enabling integration of the DML-CZ into the WDML.

The project involves designing and implementing inter-related issues in the following five areas:

1. Acquisition

- counsel's opinion of the IPR and copyright in the context of the DML-CZ
- establishing a workplace for the acquisition and preparation of printed materials
- selection of materials (journals, proceedings, monographs etc.) to be digitized
- dealing with the IPR for selected materials
- technical preparation of materials to be digitized

2. Digitization

- deciding technical parameters of digitization compatible with the WDML Best practice statements
- setting the digitization workflow
- selection and adaptation of software supporting the digitization process (existing digitization equipment within the Academy Library will be adapted and used)
- digitization of selected mathematical literature
- OCR processing and post-processing
- provision of digitizing metadata

3. Digital documents

- Digital Objects (DOs) structure specification
- defining standards for metadata (descriptive, structural, administrative – including technical and IPR ones)
- global persistent identification
- provision of complex metadata
- creation of composed DOs for various document types (journal, paper, monograph)
- archiving and presentation formats
- conversions between formats and generation of digital derivatives
- inclusion of born-digital materials
- automatic conversions of visually marked OCR data into logically structured documents

4. Digital library

- implementation of the Content Management System for the DML-CZ
- providing access to the digitized material
- interlinking of the DML-CZ content with mathematical reference databases
- research and implementation of advanced search techniques for mathematical documents
- the DML-CZ administration (including long-term preservation of the digital content)

5. Integration of the DML-CZ in the WDML

- interoperability and incorporation of the DML-CZ in the WDML (on the basis of common standards and technologies, e.g. OAI-PMH)
- linking the DML-CZ with ZMATH and MathSciNet.

Research and development in the above listed issues will always be preceded by an investigation of the existing methods and best practice procedures. Co-operation of the DML-CZ team with mathematics digitization projects in other countries will be of great importance.

The first journal to be digitized is *the Czechoslovak Mathematical Journal* (published by Mathematical Institute of the Czech Academy of Sciences). The testbed for the DML-CZ will be built upon digitized CMJ documents. The electronic material created within the DIEPER project at the University of Göttingen (digitized Czech journals *Mathematica Bohemica* and *Commentationes Mathematicae Universitatis Carolinae*) offers another possibility.

Proposed Schedule

The first year

- collection and studying of the state-of-art knowledge and trends
- establishing work contacts with other teams involved in building the WDML
- identification of relevant literature for digitization
- investigation of the IPR and copyright issues
- establishment of the workflow for metadata creation, digitization and archiving the raw digital material
- formulation of the pilot project for verifying results
- acquisition of technical equipment (dedicated server, storage capacity) for the DML-CZ.

The second year

- digitization of the material for the pilot project
- development of the prototype software system for the DML-CZ
- collecting information on existing born-digital material and investigating the possibilities for universal, efficient tools and procedures for including this literature in the DML-CZ.

The following years

- full-scale digitization
- studying tools for advanced search techniques in mathematics
- interconnecting the DML-CZ with databases ZMATH and MathSciNet
- studying interlinking documents through electronic links, unique identification and structuring documents
- preparation of organizational and legal conditions for maximizing the usability of the DML-CZ and its further development.

Partners and Responsibilities

1. Mathematical Institute of the Academy of Sciences of the Czech Republic, Prague
- project co-ordination, selection and preparation of materials for digitization, IPR and copyright issues
- *principal investigator*: Jiří Rákosník

2. Institute of Computer Science, Masaryk University, Brno
- technical integration, development of the digital library for the DML-CZ, metadata provision coordination, incorporation of the DML-CZ into the WDML
- *team leader*: Miroslav Bartošek

3. Faculty of Computer Science, Masaryk University, Brno
- OCR post-processing, techniques for searching and presenting digital documents
- *team leader*: Petr Sojka

4. Faculty of Mathematics and Physics, Charles University, Prague
- user requirements, metadata specifications, links to ZMATH and MathSciNet
- *team leader*: Oldřich Ulrych

5. Library of the Academy of Sciences of the Czech Republic, Prague
- digitization, OCR, storage and presentation of digitized content within the Kramerius system (the digital library of Academy of Sciences)
- *team leader*: Martin Lhoták

Contact

Mathematical Institute AS CR
DML-CZ (Jiří Rákosník)
Žitná 25
115 67 Praha
Czech Republic

Tel.: +420 221 403 446
Fax: +420 221 403 523
E-mail: rakosnik@math.cas.cz