

A CONTENT MANAGEMENT SYSTEM FOR ELECTRONIC THESES

Oleg Burlaca, Constantin Gaidric, Svetlana Cojocaru

The Institute of Mathematics and Computer Science of Academy of Sciences of Moldova
oburlaca@neonet.md

Abstract: PhD theses represent a global source of information that is underutilized due to their lack of physical availability. Making this information available on the Web provides the key to dissemination of the research results. This article describes ongoing efforts to build a content management system for PhD theses, authors, opponents, institutions and a web site as it's frontend. The purpose of the system is a consistent, up to date source of information and statistics for research community.

Keywords: electronic publications, theses, content management systems.

INTRODUCTION

According to an ETD (Electronic Theses and Dissertations) workshop, organized by UNESCO in september 1999, the main problems in accessing scientific research results are: limited access through traditional publishing enterprises; cost escalation for the acquisition of printed material; lack of coverage of details of research; and inadequate preservation and archiving measures for traditional (paper) material. IT techniques must be fully exploited to enhance access to scientific information, particularly to research results published in dissertations and theses.

These issues tries to address the Networked Digital Library of Theses and Dissertations [1] (NDLTD) - an international organization dedicated to promoting the adoption, creation, use, dissemination and preservation of electronic analogues to the traditional paper-based theses and dissertations. The mission of the NDLTD is to improve graduate education by developing accessible digital libraries, to increase the availability of research, and to preserve it electronically. A very promising project is NDLTD Union Catalog [2] that is a repository of rich graduate educational material contributed by a number of member institutions worldwide.

One benefit of ETDs, as mentioned by Edward Fox in the ETD Guide [3], is a reduction in the needless repetition of investigations that are carried out because people are unaware of the findings of other students who have completed a thesis. With a globally accessible collection of ETDs, students can quickly search for works related to their interest from anywhere in the world, examine and learn from those studies.

With the advent of ETDs, authors and their theses will become more visible, they will tend to produce higher quality work, and faculty will demand better writing and clearer presentation of results, if the audience for a work will increase, as opposed to the common current situation where a only a few will read the document.

But there are people that express skepticism with regard to putting theses online, providing the plagiarism as an argument. In short, many people tend to think that because a digitized thesis is easily copied in part or in whole, it can be easily plagiarized. That's why they think that it is better to keep theses offline.

The paradox of placing theses on line, especially if these theses are indexed full text can help identify analogous texts rather easily. As a result, far from placing the digitized theses at risk, putting them on line in a manner that optimizes their access, irretrievability and, therefore, visibility, offers a very efficient way to protect intellectual property and prevent plagiarism.

Unfortunately, there are supervisors that stand up against online theses. Knowing that a doctorand's thesis is unlikely ever to be read by more than a handful of people, might make a supervisor less diligent in providing supervision than is in the best interests of the doctorand.

The sharing of research results through ETDs is one of the fastest way for people working in developing countries to become known and have an impact on the advancement of knowledge.

All ETD related projects that were investigated had focused more on automatic and distributed submission of theses, library processing, full text indexing, format conversion, rather than investing in system functionality, usability, information architecture, what kind of information will be valuable in addition to a simple searchable collection of thesis pdf files. Besides the thesis itself, there are a lot of metadata associated with it: the supervisor, committee members, date of defense, the institution and department where the thesis was created and so on. This data can be stored separately from the thesis as independent units in form of related items. Such architecture can better answer questions like: the number of doctorands that a supervisor prepared; how many theses, prepared at a given institution were defended last year; which are the most popular PhD specialities; how many times a person was a committee member etc. Such a system will be more than an e-library, it will be also a valuable tool for decision support.

RELATED WORK

ETD Software

ETD-db [8] is the most widespread E-theses package in use, in part due to the support it has from the NDLTD. As of March 2004 development of the official release of this package paused at version 1.8.

DSpace [9] has been developed in partnership between HP and MIT. DSpace captures, stores, indexes, preserves and redistributes an organization's research material in digital formats. It is more mature than ETD-db and offers more functionality.

EPrints [10] creates a web based archive and database of scholarly output and is intended for use by a university or university department.

ETD Projects

Using the above software, many universities started their own online ETD library. But we searched for country-wide projects, hoping that it will have more facilities and a more consistent architecture than a separate university project. Examples of such projects are: Australian Digital Theses Program (ADT) [4], Theses Canada Portal [5], Cybertheses [6], Swedish Digital Scientific Archive (DiVA) [7].

Cybertheses, in our opinion, has a richer set of possibilities than other investigated projects. It provides an efficient indexation system and rapid searching, even while significantly increasing the visibility and the distribution of the theses. Besides the advanced search capabilities, it offers a comfortable interface for navigating a these, without the need to download it entirely as in other projects.

All ETD projects are focused only on thesis archiving, searching and distribution. There are no report modules that can help decision makers by providing statistical information.

A CMS for ETD

The reviewed ETD projects and software solutions behind them are based on individual efforts from thesis authors that submit their works to the system. The systems are designed to work without much intervention from the system(site) administrator and doesn't need handy interfaces for content management.

Our scope is to build an ETD system powered by a CMS, that will facilitate further development and enhancement of the services provided. The NeoSite CMS [11, 12] is the starting point for a CMS designed for ETD management. NeoSite was already used to develop and maintain the web site of the High Attestation and Accreditation Committee of Moldova [12]. The well structured database of institutions, specialities, scientific councils will be expanded with doctorands that pursue the PhD degree, doctor and doctor habilitat database, and of course the thesis database.

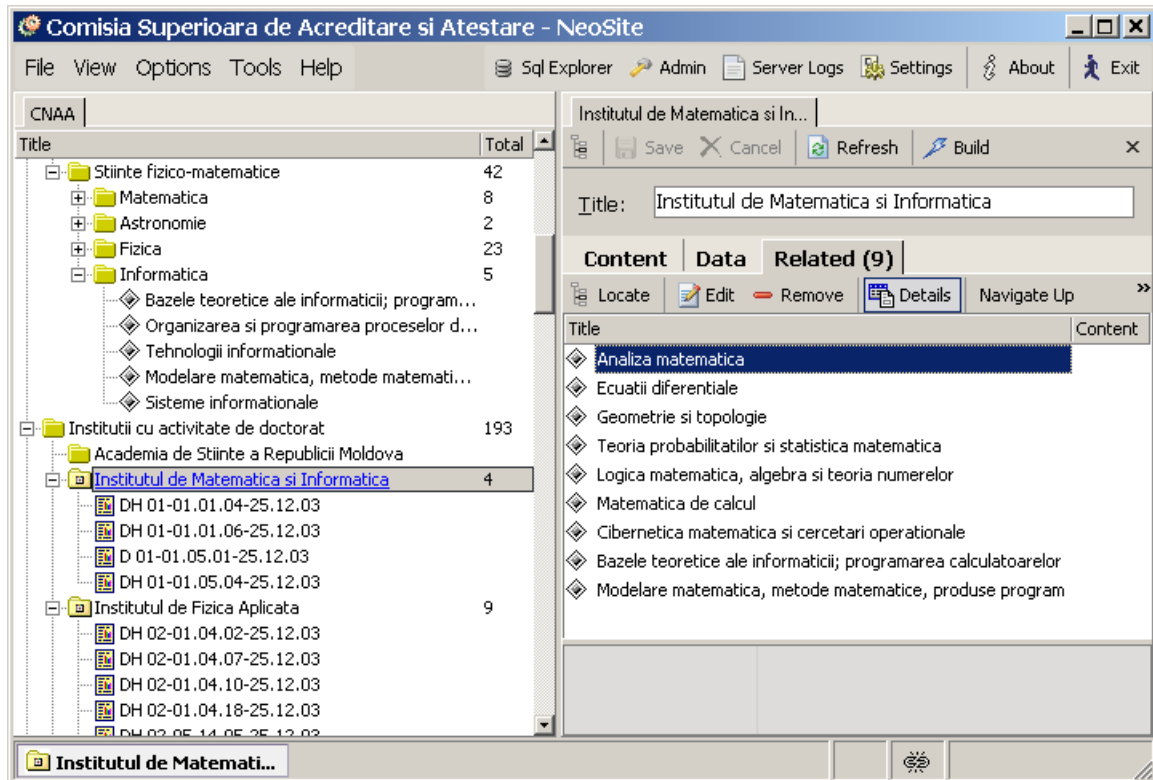


Fig.1 The CMS for High Attestation and Accreditation Committee of Moldova web site.

As stated in Introduction and shown in Fig.1, the high granularity of information objects (doctorand, thesis, institution) and the relationships between them, facilitates the development of a consistent, well structured web site based on information architecture principles. Thesis, doctorand and scientist databases will be integrated as a new layer of information, new relations will be identified and exploited between existing and new data. A new web site will be created as an interface to this database. Full text search and powerful filtering capabilities will be developed to enhance navigation.

CONCLUSIONS

Dissertations will matter more than they have in the past. Thanks to digital libraries, students and universities will pay greater attention to the quality of students research and writing.

For students, the electronic dissertation can be easier to prepare, more error free, less expensive, and more flexible in format. It can also allow more creativity on the part of the author by permitting inclusion of hypertext links and, soon, digital audio and video recordings.

Having a centralized thesis repository opens the possibility of integration with global archiving facilities like NDLTD. As stated in [13]: “Archiving should be a core activity of the NDLTD with a goal to provide a copy of each member institution’s ETD collection through collaborative persistent mirroring”.

Our work is a step towards the integration of scientific and educational institutions, research and development collectives of Moldova into the worldwide informational field. One of our missions is to increase the visibility and to valorize Moldavian scientific production at the international level.

REFERENCES

- [1] Networked Digital Library of Theses and Dissertations (NDLTD), <http://www.ndltd.org>
- [2] NDLTD Union Catalog, <http://www.vtls.com/ndltd/>
- [3] The Guide for Electronic Theses and Dissertations, <http://etdguide.org/etdguide.pdf>
- [4] Australian Digital Theses Program, <http://adt.caul.edu.au/>
- [5] Theses Canada Portal, <http://www.collectionscanada.ca/thesescanada/index-e.html>
- [6] Cybertheses, www.cybertheses.org
- [7] Swedish Digital Scientific Archive, <http://www.diva-portal.org/>
- [8] VT ETD-db, <http://scholar.lib.vt.edu/ETD-db/>
- [9] DSpace, <http://www.dspace.org/>
- [10] EPrints, <http://www.eprints.org/>
- [11] Oleg Burlaca, "NeoSite: A simple Content Management System", Computer Science Journal of Moldova, vol. 12, n. 1(34), 2004
- [12] NeoSite CMS, <http://cms.neonet.md>
- [13] High Attestation Committee of Moldova, <http://csa.acad.md>
- [14] Gail McMillan, "The NDLTD and Issues of Long Term Preservation and Archiving", ETD 2003, Humboldt University, Berlin, Germany, May 2003.