

The Integration of Lithuanian ETD into the National Digital Library

Arunas Franckevicius

IT Implementation Centre, Kaunas University of Technology, Lithuania,
arunas.franckevicius@internet.ktu.lt

Vilius Kuciukas

IT Implementation Centre, Kaunas University of Technology, Lithuania,
vilius.kuciukas@internet.ktu.lt

Antanas Streimikis

IT Implementation Centre, Kaunas University of Technology, Lithuania
antanas.streimikis@internet.ktu.lt

Aleksandras Targamadze

Dean of the Informatics Faculty, Kaunas University of Technology,
Lithuania, aleksandras.targamadze@internet.ktu.lt

ABSTRACT

Typical digital library system does not help much to get instantly online a full text of the multimedia document and other services related to processing of the retrieved document. This is the main reason why the Lithuanian Science and Studies Electronic Documents Digital Library (eLABa) project is currently being implemented. Almost all libraries hosted by Lithuanian academic institutions take part in this project. In parallel, the pilot Lithuanian ETD project funded by UNESCO has enabled to set up the repository of Lithuanian ETDs. Consequently, the main purpose of this paper is to present how the ETD repository is going to be redeveloped and integrated into eLABa.

To create an experimental Lithuanian ETD library as part of the eLABa the Fedora 2.1 software and complementary VTLS product VALET for ETDs has been chosen. In particular, the advantages and drawbacks of using Fedora 2.1 based repository along with the VTLS will be discussed. In general, ETD repository serves as a good model for creation of the eLABa digital library.

Keywords: ETD, metadata, e-publishing, repository, digital library

1. INTRODUCTION

Lithuanian Academic Libraries Network (LABT, 2006) cleared the way for creation of the Lithuanian Virtual Library with unified access to digital resources as well as to the other typical library services. The conceptual LABT model has been developed (Kuciukas, 2004) taking into account many factors including both advancements in information technology and increased needs of users. In accordance with the conceptual model the electronic library of Master's theses, Doctor's dissertations and their summaries (Streimikis, 2005) has been developed. For that the VTech ETD software system was adapted and used. The realization of the Lithuanian ETD library made it possible firstly to experiment and to test in practice the fundamental principles of the conceptual model, and secondly to extend considerably the research and development activities in creation of Lithuanian Electronic Academic Library (eLABa). The eLABa is viewed as databases of full-text publications distributed across the network of academic libraries and, importantly, includes necessary e-publication facilities and services. Three projects supported by European Union Structural Funds (ES SF) currently are set up aiming at creation of the academic e-publishing system.

The paper describes briefly the overall goals of the eLABa, presents the collections of e-documents supposed to be stored in the library, on the ground of Lithuanian ETD deployment discusses strengths and weaknesses of VTLS software VALET for ETDs, especially from the perspective of integration with database management system Fedora, which was chosen as a base platform for eLABa repository.

The main goal LABT project is to create the Lithuanian Virtual Library through the automation of academic libraries, unification of search and access to electronic (hereinafter – e) information sources and virtual services. In 2002-2004 considering the latest trends in information technology and the increased needs of users the conceptual LABT model (Kuciukas, 2004) providing the conception of Lithuanian Virtual Library has been prepared. Following this conception in 2003-2004 the electronic library of Master's theses, Doctor's dissertations and their summaries (Streimikis, 2005) based on the adapted VTech ETD software has been created. Lithuanian ETD library allowed checking in practice the fundamental principles of the suggested general model for Lithuanian academic e-publishing, full-text databases and Lithuanian Virtual Library.

Three projects related to the academic e-publishing (“The creation of Lithuanian electronic publishing system”, “Lithuanian science and study e-documents storage and presentation to the users” and “Lithuanian virtual library development and creation of full-text database repository”) financed by ES SF being implemented in 2005-2008 is the natural outcome of the LABT project which has been carried out within eight years.

On the basis of LABT and above mentioned three projects financed by ES SF was started the development of the Lithuanian information system of science and studies electronic documents. Actually this system is being created as Lithuanian Electronic Academic Library (eLABa) functioning as a national digital library. Fedora software has been chosen as the base of eLABa repository. In the article are discussed the goal of eLABa, the collections of e-documents supposed to be stored, the usage possibilities, advantages and problems of VTLS software VALET for ETDs, operating with Fedora met in the process of integration of Lithuanian ETD library into eLABa.

2. GOALS

The goal of eLABa is to create the environment and means necessary for preparation, collection and long term preservation of the Lithuanian science and studies electronic documents grouped into various collections and to deliver them to the internet users in different outputs.

The main functions of eLABa:

- submission of science and studies publications into eLABa using Lithuanian e-publishing process workflow models;
- gathering, filing, accumulation and long term preservation of science and studies electronic documents;
- information search and the direct and constant access to the science and studies documents stored in eLABa under the conditions set by authors or copyright subjects for the employees and students of Lithuanian science and studies institutions, and other users;
- submission of data about the electronic documents of science and studies to NDLTD, Google and other information search systems.

The eLABa project is aiming creation of conditions for faster dissemination of up to date knowledge and material for studies by using information and communication technologies, improvement of the quality of scientific research and studies, development of distance learning and whole-life education, promotion of the activities and achievements of Lithuanian scientists and science and studies institutions, education and development of knowledge society.

3. COLLECTIONS OF E-DOCUMENTS

Classification of the resources according to the documents types (books, journals, ETDs, etc.) facilitates access and retrieval of knowledge and material for studies more effectively. At the first stage (starting from 2006) five collections of e-documents are planned to be accumulated in eLABa.

The Collection of Books. Science, educational and information publications (books), for example, monographs, collections of articles, course books, educational books, learning books, task books, learning programs, summaries, reference books, dictionaries, albums, encyclopaedias of various themes published by Lithuanian publishers are assigned to the collection of books. The search for a book can be performed using search indexes such as author, title (name), ISBN, publisher, place and year of publication and etc. The search of contents can be carried out using key words (terms or words from contents of the book, annotations and summaries). All books are grouped to thematic categories.

The Collection of Journals and Articles. The scientific journals of Lithuania and the articles of scientists of Lithuania published in various foreign scientific journals are assigned to the collection of journals and articles. The necessary journal can be searched according title, ISSN, publisher and etc. Articles can be found browsing the numbers of chosen journals or searching by the key words (author of the article, title, terms or words published in the summary of the article). Journals and articles are grouped to thematic categories according the branches and trends of science.

The Collection of Scientific Reports. The reports of contracted scientific research carried out in Lithuania are assigned to the Collection of Scientific Reports. Scientific reports are grouped agreeably to the thematic categories. The scientific reports can be found by title, by name of scientific coordinator, by the institution name, by registration number, place and year of publication, etc. The search of contents can be carried out using keywords (terms or words which are published in the summary of scientific research).

The Collection of Events Papers. The papers, the handout of the papers, the presentations of the papers of conferences, seminars, congresses and other events held in Lithuania, and the papers of Lithuanian scientists published in proceedings of international conferences are assigned to the Collection of Events Papers. The necessary publication of event papers can be searched according the title of event the title of publication or paper, ISSN, ISBN, publisher, place, year of publication and etc. The papers of event one can find browsing the publications of chosen conferences or searching using key words (author of the paper, title, terms, or words published in the summary of the paper). The papers of the events are grouped according to thematic categories.

The Collection of ETD. Theses, their summaries and master's works defended for a degree in institutions of science and education of Lithuania are submitted to ETD collection. The necessary theses, their summary or master's work can be searched according author, title (name), institution granting a degree, supervisor, place, year of publication and etc. The search of contents can be carried out using key words. Theses, summaries and master's works are grouped to thematic categories according the branches and trends of science.

Collection of ETD, Collection of books, Collection of journals, Collection of scientific reports and Collection of events papers are open for all users. Full texts of e-documents of these Collections are usually presented in PDF but it can be also presented in other formats, for example XML.

In the near future eLABa is supposed to be supplemented with other collections of e-documents such as a collection of music compositions, a collection of empirical data, etc.

In the next sections the eLABa component for the formation of ETD collection (eLABa ETD) is described in more detail.

4. ABOUT eLABa ETD

Seeking to specify more clearly the eLABa ETD component, which is being developed it would be reasonable to specify its modular structure and possibilities of using VALET for ETDs for the development of this component. Experimentally identified functional and technical limitations of VALET for ETDs interfering the development of eLABa ETD component are disclosed and recommendations for the elimination of these limitations are given.

4.1. The Modular Structure of the eLABa ETD

Three layers can be highlighted in the modular structure of the eLABa ETD: Frontend layer, Integration layer and Pluggable layer (Picture 1).

4.1.1. Frontend Layer Modules

Submission module is intended for the submission of theses, dissertations and their summaries (hereinafter - ETD) to eLABa repositories. This module is being used by postgraduates and doctorate students (hereinafter – students). Using submission module it is possible to submit the whole ETD or its metadata. Also by using this module the primary verification of metadata is performed.

Review module is intended for revising, reviewing, validating, approval and submission of ETD and its metadata to eLABa repositories if no mistakes are found. Reviewing is carried out step by step in several stages and if needed by coming back to the previous stages. After the reviewing is finished ETD and its metadata are submitted to eLABa repositories. This module guarantees the connection between a reviewer and a student submitting his/her ETD for example by using e-messages. A reviewer and a student intercommunicate using this module in order to correct as operatively as possible the mistakes that have emerged in ETD or in its metadata.

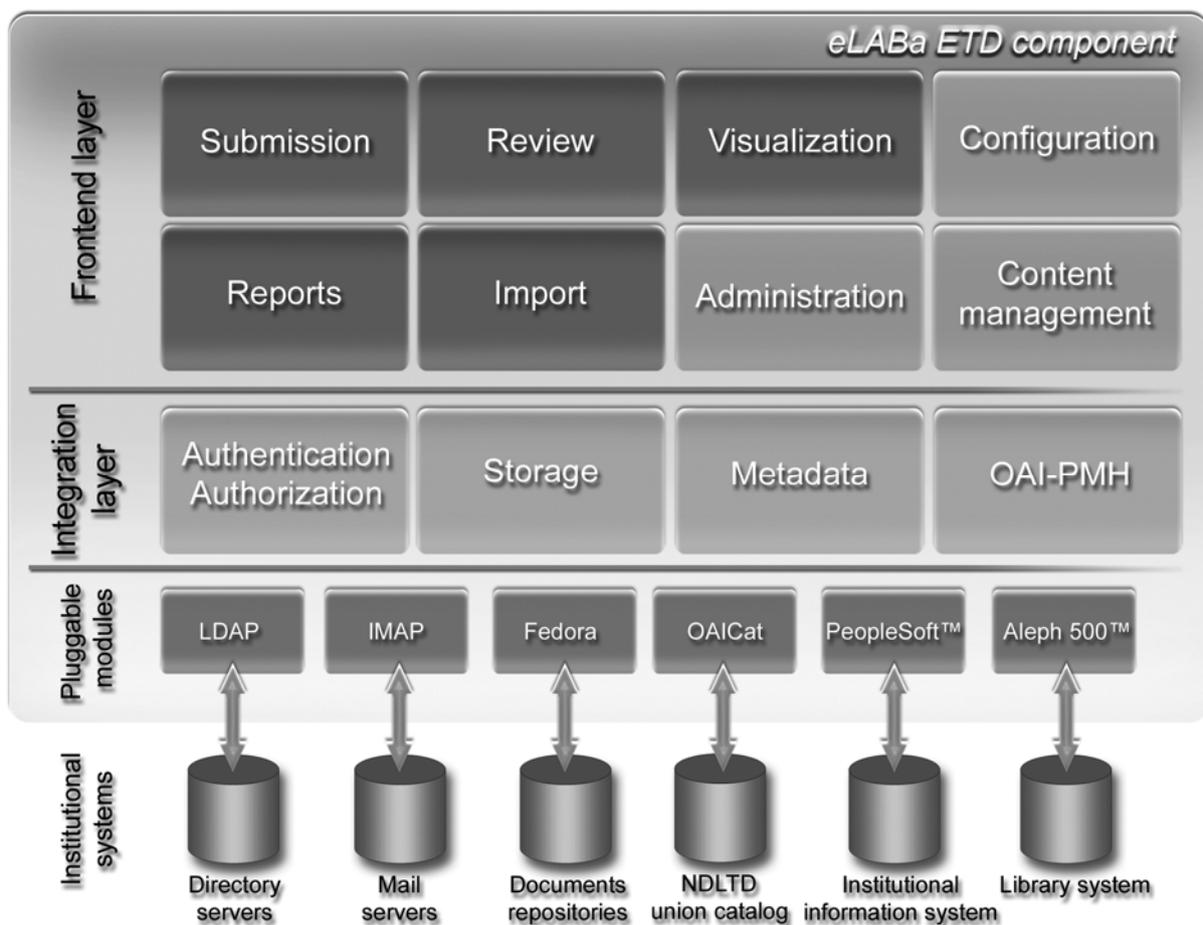
Reports module is intended for generating the various reports about ETD stored and submitted to eLABa and for accumulating all required statistical information such as the types of the submitted ETD, their limitations, etc.

Import module is intended for the processing of specially prepared ETD by selecting metadata from ETD according to the corresponding template automatically, and later verifying and transferring them to the review module for processing.

Visualization module is intended for the correct visualization of ETD stored in eLABa repositories in different sections. The module must ensure not only the visualization of the stored ETD in different e-formats that meet the requirements of the users but also must provide the possibility to navigate through the tree of document structure for example through sections, paragraphs, etc.

Administration module is intended for user administration functions such as assignment of rights to a particular user or to the group of users, users list management, etc. used for the administration of the main eLABa ETD users (students, reviewers and other users) and their groups (so called CRUD – Create, Retrieve, Update, Delete functions).

Configuration module is intended for the description of various eLABa ETD interfaces and for configuring its system parameters. For example the interface for submission and reviewing Master's theses should differ from the interface for submission and reviewing Doctor's dissertation. It is reasonable to design different system interfaces for different institutions, their faculties, etc. Using this module system eLABa ETD parameters should also be configured for example various registers with different values depending on the configured interface.



Picture 1. The Modular Scheme of eLABa Being Developed

Content management module is intended for revision and administration (CRUD functions) of both ETD metadata submitted to eLABa repositories and administrative data, also for the export of ETD and their metadata in various formats such as XML, PDF, PS, etc.

4.1.2. Integration Layer Modules

Authentication Authorization module is intended for user authentication according to its identity for example by its name or unique student card ID and password, also to identify if the user is authorized to use one or another hardware and software resources. This module is common for both students submitting their works and reviewers and may be integrated with different external systems such as LDAP, IMAP, Active Directory, etc.

Very important function of this module would be to create and assign access rules (ACL – Access Control List) to e-documents to be stored into eLABa repository following by author copyright license.

Storage module is intended to perform and manage the temporary and long-term preservation of ETD and accompanying metadata in various repositories. After choosing a proper control mechanism this module must guarantee the possibility to operate not only with Fedora but also with other repositories such as DSpace, Greenstone and others if needed, thus ensuring independence from a particular repository.

Metadata module is intended for the collection of data related to ETD from information systems (such as PeopleSoft) of various institutions as well as library systems, thus reducing the number of mistakes of metadata formation and acceleration the ETD submission process. The data gathered during this process are sent through the submission module to a student, who reviews, corrects,

supplements them and approves their correctness. This module is used to transform the ETD related data submitted to eLABa repositories into MARC bibliographic records, to load them into Lithuanian ETD union catalogue and/or into library systems (such as ALEPH 500) of institutions by indexing them. Important functionality of this module – ability of automatic metadata format transformations such as DC, MARC21, UNIMARC, etc.

OAI-PMH module must enable other external harvesting systems (such as OAICat) which forms NDLTD union catalogue to take ETD metadata in OAI-MPH protocol. In such a way the search of ETD according to various search criteria such as author, work title, year of creation, topic, etc. is ensured for external search systems. Such search may be performed through the portals of library systems of institutions, through the portal of Lithuanian Virtual Library and other systems such as NDLTD, OAIster, Google and other portals.

4.1.3. Pluggable Modules Layer

This layer is used to ensure the connection of eLABa ETD with the external institutional systems for example with user databases (LDAP server), e-mail systems (Mail Server), repositories (Fedora, DSpace, Greenstone), database management systems (ORACLE, Versant), library and other systems related to the search of e-resources (ALEPH 500, MetaLib, SFX). This may be guaranteed by creating the corresponding interfaces used in the integration layer and developing new so called switching modules. For each separately used external realisation of the system the separate switching module hiding the external system from eLABa ETD is required.

4.2. Functional and Technical Limitations of VALET for ETDs

While planning and implementing eLABa ETD, which main modules are described in the section 4.1. it has been decided to examine in-depth the possibilities allowing the integration of VTLS company's product, VALET for ETDs, into eLABa ETD. After carrying out the experimental research the functional and technical limitations interfering the integration of VALET for ETDs into eLABa ETD were identified and the ways to eliminate the limitations were found.

4.2.1. Functional and Technical Limitations of VALET for ETDs and Ways for their Elimination

Security and authentication/authorization. VALET for ETDs allows integrating the different authentication systems such as IMAP, LDAP, Active Directory and others. However actually it performs only the function of user identification, i.e. authentication function, which uses only the one selected authentication system and do not have the possibility of managing user groups and rights. Also VALET for ETDs have no possibilities which could facilitate the import of users' authentication/authorization data from external systems. For example there is no possibility to import student's name, surname, faculty, etc. automatically from LDAP servers of Lithuanian institutions participating in ETD project (hereinafter – institutions) into VALET for ETDs.

Seeking to secure the listed lacking possibilities, a new authentication/authorization module described in the paragraph 4.1.2 is recommended to be developed. VALET for ETDs program code would be enhanced in a way to give possibility of user and users group access to repository resources authorization using ACL. SSO (Single Sign-On) functionality would be added to VALET for ETDs to give possibility central user authentication and authorization systems as to be used by group of software systems: eLABa, ALEPH 500, MetaLib, etc.

The import of metadata from external IS. Many of Lithuanian institutions have their own information systems (IS) where additional ETD metadata such as information about ETD supervisors, dissertation defence commissions, reviewers, etc. are stored. The mentioned information is also stored in Lithuanian science and studies information system (LieMSIS) which is being implemented now in

all Lithuanian science and studies institutions. On purpose to avoid routine work and to reduce the probability of mistakes entering the additional metadata it is reasonable to import this information automatically from LieMSIS and institutional IS.

But VALET for ETDs has no possibility to integrate external IS and cannot import ETD related into any widely used format such as plain text, MS Word, Excel. Therefore it is recommended to enhance Submission module of VALET for ETDs with new functions which would enable to obtain the needed newest additional data from the corresponding institutional IS during the process of ETD submission. The new functions are recommended to be implemented in such a way that they could be easily adapted for new institutional IS.

Postgraduate theses by several authors. In many of Lithuanian universities it is practised that one Master's thesis can be written not by one but by two or three postgraduates. But according to the logic of VALET for ETDs the ETD prepared by only one author can be submitted to the repository. Therefore Submission module and Review module of VALET for ETDs are recommended to be supplemented with new features which would allow entering new metadata describing ETD, for example co-authors.

Reports. The essential constituent of a modern IS is the formation of various reports. Therefore for creation of ETD collection the possibility to form reports according to various features for example according to ETD types, users, institutions, etc. is essential. However VALET for ETDs does not have the possibility of forming reports.

In order to guarantee the possibility of eLABa ETD reports, a new Reports module described in the paragraph 4.1.1. is recommended to be developed.

Configuring. The practice has shown that the architecture of VALET for ETDs based on so called logical units of base configuration (called View), does not allow easy making of many required changes in configuration system for example forming a new submission and reviewing environments, setting a few languages in the submission interface, etc.

In order to guarantee the flexible configuring the architecture of VALET for ETDs is recommended it to be basically improved as nonessential changes are not sufficient. For example in order to install multilingual feature in VALET for ETDs operating in one View, the essential changes must be carried out in many modules applying them to the classifiers and templates of different languages which could be easily and promptly interchanged.

Administration. VALET for ETDs does not have an administration module which simplifies the administration of the system. Therefore VALET for ETDs can be administered only using configuration files. The practice has shown that such administration of VALET for ETDs is inconvenient and inefficient.

Seeking to guarantee a convenient eLABa ETD administration, a new User administration module described in the paragraph 4.1.2. is recommended to be developed.

4.2.2. Technical Limitations

Perfection. The current version of VALET for ETDs only partly meets the needs of eLABa ETD Submission, Review and Storage modules. The main disadvantages of these modules are described in the paragraph 4.2.1. The practice has shown that if Submission, Review and Storage modules are improved and integrated with the needed new Authentication/authorization, Reports, Import and Metadata gathering modules, in principle VALET for ETDs could be used for the development of eLABa ETD.

Architecture. VALET for ETDs is professionally developed. But the chosen architecture does not ensure an easy appliance of VALET for ETDs to the specific needs. It's modules directly depend on each other. Changes in one module require changes in other modules. The flexible possibility of integrating other external systems, except User Authentication, to VALET for ETDs are not foreseen.

Therefore it is reasonable to make essential architectural changes in the newer versions of VALET for ETDs following the established principles of OSA (Open System Architecture).

Effectiveness. VALET for ETDs runs under CGI (<http://hoohoo.ncsa.uiuc.edu/cgi/>). This is an outdated WEB application interface. Running such application a separate process is created for each query by using additional technical and time resources. Also CGI interface is inapplicable in case the number of parallel user queries is high. Therefore WEB application interface is not so effective as Java2EE (<http://java.sun.com/javaee/>) application servers for example. Therefore seeking for better effectiveness, VALET for ETDs should be developed in another language such as JAVA and this requires too much time and other resources. Therefore the effectiveness of VALET for ETDs is recommended to be increased by using Apache WEB Server extensions such as mod_perl for example.

Migration. Although VALET for ETDs has been developed using Perl programming language due to some specific limitations used for the management of UNIX command files for example, it runs only under UNIX environment. Thus the transfer of VALET for ETDs into MS Windows environment would be rather complicated.

Therefore in new VALET for ETDs versions the migration problem is recommended to be solved.

Maintaining and development. Perl programming language used to develop VALET for ETDs is rather hardly readable. Therefore it is reasonable to improve VALET for ETDs modules as little as possible. Seeking to implement the required features which are lacking in VALET for ETDs, new independent modules are recommended to be developed. It is purposeful for VALET for ETDs to have external interfaces (e.g based on SOAP/REST), enabling to enhance/change its functionality by using other systems. Later it is recommended to integrate the improved VALET for ETDs and newly developed modules into a separate component. The component formed in such a way could be easily updated in the future for example by replacing the current VALET for ETDs with a new version. Such principle of a separate component is applied in the development of eLABa ETD.

Documentation. VALET for ETDs is poorly documented. Therefore it is recommended to complement VALET for ETDs documentation substantially.

5. CONCLUSIONS

In 2005-2008 on the basis of LABT and three projects financed by ES SF the Lithuanian Information System of Science and Studies Electronic Documents was started to be developed. Actually this system is being created as Lithuanian Electronic Academic Library (eLABa) functioning as a national digital library. As the base for eLABa repository Fedora software has been chosen. At the first stage five collections of e-documents, including ETD collection, are planned to be accumulated in eLABa repositories.

After analyzing the needs of eLABa ETD component its modular scheme has been prepared and functional requirements of modules have been formed. Seeking to reduce the realization costs it was decided to use OSS (Open Source Systems). Currently most popular OSS for the formation of ETD collection is VALET for ETDs software, which can be integrated with Fedora. Therefore the architecture and functional possibilities of the VALET for ETDs have been analyzed in more detail.

The practice has shown that current features of VALET for ETDs are insufficient for the creation of the needed eLABa ETD component. VALET for ETDs only partly meets functional requirements of

three (Submission, Review and Storage) modules from twelve described in the article. Therefore VALET for ETDs must be substantially improved in order to integrate it into eLABa.

For the formation of the aimed eLABa ETD component it is reasonable to improve all modules of VALET for ETDs and to develop new independent modules which would realize the required possibilities. For the development of new modules it is reasonable to choose such architecture, which would secure their operation irrespective of VALET for ETDs version. It would allow changing VALET for ETDs with a new VALET for ETDs version or with entirely another system meeting better eLABa ETD needs.

After making the first tests with eLABa ETD prototype developed on the base of VALET for ETDs it has been determined that this prototype could be started to be applied for the submission of ETD and their metadata to eLABa repositories. It would be reasonable to continue the development of this prototype with a perspective to create the aimed eLABa ETD component which would secure all its functions required.

6. REFERENCES

Kuciukas, V., Abramcikiene, R., Butkeviciene, B., Duobiniene, G., Jauckojis, J., Jurgutis, et al. (2004). *Feasibility study of the Lithuanian academic e-publishing*. Retrieved March 29, 2006, from Kaunas University of Technology, Lithuanian Academic Libraries Network Web site: <http://www.labt.lt/naujienos/E-leidybos-VDDDB-VB-poreikiu-ir-galimybiu-studija.pdf>.

Lithuanian Academic Libraries Network (LABT) (2006). Retrieved March 29, 2006, Kaunas University of Technology, Lithuanian Academic Libraries Network Web site: <http://www.labt.lt/en>.

Streimikis, Antanas, Kuciukas, Vilius, & Targamadze, Aleksandras (2005). Lithuanian ETD Project ETD2005. In *Evolution through Discovery: 8th International Symposium on Electronic Theses & Dissertations*, 28-30 September, 2005, Sydney, Australia [pp. 1-6]. Retrieved March 29, 2006, from Web site: <http://adt.caul.edu.au/etd2005/papers/127Streimikis.pdf>