Available online at: https://ijact.in

An International Journal of Advanced Computer Technology

CHALLENGES IN CREATING UNIVERSITY DIGITAL DOCUMENT REPOSITORIES

Emil Hadzhikolev[1], Stanka Hadzhikoleva[2], Nedelcho Andonov[3]

[1, 2, 3]Plovdiv University "Paisii Hilendarski", Plovdiv, Bulgaria
[1]emil.hadjikolev@gmail.com, [2]stankah@gmail.com, [3]dechosl@gmail.com

Abstract: The educational institutions in the system of tertiary education are business organizations with an extremely complex structure. A great variety of versatile processes takes place in them like selection, educational, administrative, financial, accounting, etc. The participants in them have manifold and sometimes variable and/or elective roles like university students, academic staff members, employees, occupying administrative positions, deans, rectors, Academic Council members, etc. The use of centralized digital document repositories solves various document flow-related problems, expedites business processes and reduces the possibility of making mistakes. This paper presents different challenges of creating university documents repositories. It also describes a system of objects, interconnections and principles that offers a possible solution.

Keywords: university document repository; digital document repository; university repository model.

I. INTRODUCTION

Manifold business processes, related to different activities (admission of prospective students, administrative support for staff and student, administration of student grants and accommodation, human resources management, academic staff development, quality management, accounting, etc.), are being carried out in educational institutions in the system of higher education. Some of these activities are managed centrally - at a university level, others are administered at a faculty level. In this process a great number of various documents are being created, that are further administered by the manifold organizational units and exploited by different types of users – software applications, staff, Academic Council members, students, etc.

Lots of organizations use legacy document systems by means of which they partially manage and archive the document flow. A real risk exists of storing an expired version of a document for the purposes of operational work by different departments and organizational units. In some cases, each department stores its own documents and places them at the disposal of other units on request. A natural solution to the problem is the implementation of a centralized document repository with different levels of access depending on the type of job responsibilities of the different staff members.

This paper presents a certain approach to creating a university document repository. It is focused on the possibility for detailed modeling of the basic objects and subjects of the repository. The strict typification of objects is intended to control the personalized user access to the resources stored in the repository, facilitate the searching for and accessing the documents through web-based services and the integration with other institutional applications.

II. DIGITAL REPOSITORIES AND DOCUMENT MANAGEMENT SYSTEMS

Research and experimental activities are constantly carried out in educational institutions. Various digital repositories and catalogues for the purpose of organizing and storing manifold information are used. This includes library catalogues containing information about books and textbooks [1], metadata repositories, describing catalogues of educational resources for teacher training [2], shared repositories of educational materials in a multilateral setting [3], repositories for project-based education resources [4], such, containing the results of experiments conducted in the process of education of engineers [5], specific educational resources created by students [6] or such used in the educational process [7], models of psychological surveys are being tested [8], etc. In recent years a great number of software systems enabling the storage and management of digital documents have been created. Among them we distinguish between paid and free systems, ones with an open source, others, created solely for the purpose of utilization by educational institutions, etc.

Document management systems are applications used for registering, tracking, managing and storing documents. They generally support components sustaining the processes of metadata management, data integration and validation, indexing, document security and access, work-flow management, teamwork facilities, document version control, search facilities, etc. A great variety of applications designed to meet the requirements of educational institutions are available.

DocuVantage OnDemand [9]is an application intended for centralized storage of documents. It has options allowing the editing and sharing of files with a certain user group and generates detailed information about the user who has accessed, viewed, modified and approved any kind of information. The application supports options for defining workflows which allows for the automation of document routing.

Docsvault[10]supports basic functionalities for document management and profiling, indexing and tagging of documents, etc. Users are able to create customized templates for their paper and electronic documents. Docsvault also supports options for task management, e-mail management, etc.

Treeno[11]supports electronic document routing, automatic tracking and reporting capabilities concerning the documentary flow. It optimizes the activities related to contract management, budget revisions, third-party auditing, etc. A specific functionality of the application is the automation of the process of creating project proposals that, in its turn enables the control of adhering to contract-related timelines and guarantees confidentiality.

DynaFile[12]support integrated on-line forms and e-signatures, that enable automatic data management. Administrators have full control over user access rights with the possibility to allow access only to the type of document the respective user needs. The application can be integrated with the institutional education management system thus enabling the automatic synchronization of students-related data.

Other widely used document management systems are*eFileCabinet*[13],*FileHold*[14],*Virtual Cabinet*[15] etc.

Our survey shows that document management systems used by educational institutions support a wide variety of functionalities that partially automate educational, administrative, auditing, financial and other processes.

III. MODEL OF UNIVERSITY DIGITAL DOCUMENT REPOSITORY

A. Basic Requirements to the University Document Repository

The typical applications used in universities are, as follows: software used by Student Affairs Departments, Human Resources and Quality Management software, Grants and Dormitories management applications, etc. University and Faculty websites often publish information that is automatically extracted from other software applications. The development of a centralized university application would ensure that all these applications publish and extract data directly from the digital repository, rather than by interaction with one another (fig. 1).

The basic requirements that a centralized repository must meet are:

- It has to allow external software systems to upload data to the repository, perform automated search, access and download documents;

- Multi-user support capabilities – to support various user roles, with different access rights regarding viewing, uploading, updating and deleting documents;

- Strict centralized user management control system and distributed document management control system;

- Ability to describe and maintain metadata about the documents, depending on their type;
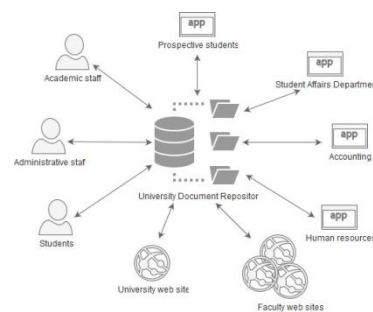
- Document version management system, etc.

Figure 1. Centralized document repository

In order to be universal, the repository has to be configurable. This implies meeting some ***additional requirements:***

- Strict typification and categorization of documents;
- Modelling of document description standards;
- Each separate document must be described by multiple metadata – some of them are obligatory, and another – optional;
- It should be possible that new document describing metadata be added at any time, without the need to change the applications using them;
- Support of document describing standards, etc.

B. *Digital Model of a Document*

To enable different software applications to search for and extract documents from the repository, metadata for each separate document must be stored, related to different aspects of the respective document - origin and purpose of the document, contents, access control, etc. For the purpose of describing a document we suggest the adoption of a 4-layer model, consisting of a basic layer, categorization layer, access control layer and metadata layer. Each layer contains a great variety of specific characteristics [16].

The integration into the university information ecosystem of new software applications, utilizing the digital repository or amendments of the university or national regulatory framework *may also lead to imposing additional requirements on document models*. In cases of insufficient system design flexibility, the integration of additional characteristics and metadata may necessitate additional programming activities – for the purpose of the visualization of new characteristics, storage of items in the database, their further processing, etc. To avoid this, it is appropriate to create a document modeling software module in which the user can determine the metadata that characterizes a particular document or category of documents. That allows for an additional level of abstraction, which would provide a universal way of documents description. For this purpose it is necessary to define some basic terms: *"standard", "document type", "and category", "operational unit"*.

A *"Standard"* is an abstract model of a document that defines a framework for the description of a class of documents. *"Document type"* is a term defining more precisely the contents and intended purpose of a certain document. Document types can be classified into different *categories*. It is possible to create one or more categorizations used for the needs of different applications, organizational units, and custom document searches. A document of a certain type may belong to multiple categorization groups. The model of a document is described in [16].

C. *Models for Application Access Control Management*

A wide variety of *access control management models* has been developed in recent years.

Role-based Access Control (RBAC) is a model based on roles. Specific roles are assigned to system users and rights are further assigned to specific roles. Thus users are not assigned permissions directly, and acquire them through their role (or roles) [17].

Attribute-based access control (ABAC) is another widespread method of access control management. Here the requests of a given subject to perform certain operations with objects are allowed or disallowed on the basis of certain subject- and object-related attributes, environment conditions and policies, specified in the terms describing these attributes and conditions [18].

Organization-based access control (OrBAC) allows the policy designer to define a security policy independently of the implementation. The chosen method of fulfilling this goal is the introduction of an abstract level where the roles, activities and views abstract the subjects, actions and the object concepts. For each multiple objects distinguished on the basis of belonging to a certain role, activity or view, the same security rule applies. Each security policy is defined for and by an organization. Thus, the specification of the security policy is completely parameterized by the organization so that it is possible to handle simultaneously several security policies associated with different organizations [19].

Role and Organization Based Access Control (ROBAC) is a model intended to upgrade OrBAC by way of defining security policies, associated to multiple organizations. Unlike the RBAC model, where the user access rights depend solely on the role assigned to him, in the case of ROBAC access rights are determined by two characteristics – the role assigned to a subject and his belonging to a certain organization. In ROBAC permissions are defined as operations carried out with object types rather than with concrete objects. A certain user can access an object solely on condition that a "role-organization" pair is associated with him, the role has the rights to access the given object type and the organization "owns the object". Different ROBAC-model variations are described in [20, 21].

The basic elements of the access control models are:

- *Objects (resources)*, access to which is being permitted – files, data, processes, assignments, etc.;
- *Subjects (users)*, accessing the objects – users (people), applications, processes, etc.;
- *Operations*, that can be carried out with the objects by the subjects – reading, saving etc.;
- *Permissions* – define the *rights* determining the *set of specific operations a certain user is allowed to perform with the objects*.

Different access control models define different *mechanisms for assigning permissions*, which can be given by an administrator and / or automatically.

D. *Objects, Interrelations and Principles of a University Repository*

In view of the limitations predetermined by the specifics of educational institutions, we propose an approach that includes the following *basic objects, interrelations and operational principles*:

1. A *single user* may be assigned *several roles* in different *operational units of the organization* (Department member, Faculty Dean, etc.).
2. *Operational units*:

- constitute structural units of the organization (Faculty of Physics, Faculty of Mathematics and Informatics, Algebra Department);
- are *organized in a hierarchical structure* - (The "University" level is divided into different faculties, affiliates, and further subdivided into departments under the respective faculties, etc.;
- provide *possibility to inherit rights in the organization structure hierarchy* – ascending and/or descending.

3. *Types of operational units*:
   - Specify the types of operational units (University, Faculty, Department, Administrative unit, etc.);
   - Predetermine the *common security policies and policies for operating with objects belonging to them* – users and documents;
   - Have specific *hierarchical structure*, that *determines the rules for valid interrelations between the different operational units* within the organization (the "University" type operational unit can only include "Faculty", "Administrative unit" and "Affiliate" type units, but not "Department" type unit);
   - Provide *possibility to inherit rights within the unit type-based hierarchical structure*-ascending and/or descending.

4. *Roles:*
   - *Determine common security and access rights* (Rector, Deans, Faculty Members etc.);
   - A certain role can be associated to *multiple users*;
   - A certain role is valid only within certain *types of structural units*;
   - A role can be *mandate-based or elective* – has a predefined validity period;
   - Different *role-based hierarchical structures* can be created for the purpose of allowing the *hereditary distribution of access rights* – a Dean (of a faculty) and a Director (of an affiliate) can hold identical rights of access to the structures under their subordination and be assigned a common "Parent" role – "Dean/Director".

5. *Documents:*
   - Can be grouped in different *document types*, with the purpose of carrying out single-type operations with them;
   - Belong to *standards*, determining *standard characteristics*;
   - Can be *grouped into categories*, which facilitates searching and other automated group-based activities performed on them.

6. *Types of documents:*
   - Predefine *common design and characteristics* of the documents of a given type;
   - Predetermine *common access rights* to all the documents of a certain type;
   - A certain *type of document* can be allocated only to some *predetermined types of operational units;*

- *Multiple of document types* can be *organized in document type groups* to facilitate the process of access rights administration.

7. *Operations* performed on document types:
   - *Basic: writing and reading;*
   - *Additional* (the realization and use of which is not obligatory):
     o *related to writing a document:* editing, publishing, confirmation, etc.;
     o *related to reading: reading of the document heading only; reading of both the heading of a document and metadata related to it*, etc.

8. ***Kind of operations. Depending on the association of a document to a certain organizational unit*** and with a purpose to simplifying and securing the process of granting access rights, operations can be divided into two groups:
   - *Private*: for users at a *Document owner-level* or *ascending and descending hierarchical levels* originating from it:
     o *basic operations*: private write, private read;
     o *additional operations*: all additional operations included on p.7 can be preset as private ones.
   - *Public* – available to all remaining users (that do not belong to the Document owner-level or ascending and descending hierarchical levels originating from it.):
     o *basic operations*: public read;
     o *additional operations*: all additional reading operations included on p.7 can be preset as public ones.

9. ***Permissions – the access rights of a certain user*** are predetermined by his *role*, *affiliation to specific operational units* (and their respective unit types) and the *type of the document* (or the respective unit and unit type to which the document is registered as available). For instance, at the "Faculty Council Decisions" document type level, a user to whom the "Dean" role is assigned can only access documents affiliated to his particular operational unit, but not such, associated to any other unit. Default policies facilitating the process of administration and enabling *default permissions* (that can be altered by administrators only):
   - *For private operations:*
     o the roles at the "Owner" level have private write and private read permissions;
     o the roles up and down the hierarchical structure have private read permissions but don't have private write permissions;
     o private permissions are not applicable to operational units not belonging to the hierarchical structure (ascending or descending) originating from a specific unit type.
   - *For public operations:*
     o users associated with roles assigned to the current type of operational unit and such

hierarchically ascending from it are granted public read permissions;

o users associated with roles assigned to the current type of operational unit and such hierarchically descending from it are not granted public read permissions.

In order to optimize the process of rights configuration, as with the ROBAC model, we use the typification of structures which enables the simultaneous granting of permissions to a group of objects. Access to the repository is being configured through the several basic steps:

1. *Modelling of hierarchical structure of types of operational units;*
2. *Modelling of roles,* corresponding to the exiting academic and administrative positions;
3. *Configuring the affiliation of roles to operational unit types;*
4. *Modelling of document-related standards;*
5. *Modelling of document types;*
6. *Modelling of permissions related to specific document types.*

All these enable the step-by-step configuration of the repository.

IV. CONCLUSION

Educational institutions have a complex structure, within which a wide range of business processes are conducted. Existing document management systems specialized in use in higher education institutions support partially different sets of business processes and cannot fully meet their needs.The development of digital document repository, that permits multiple-user and application access to the documents stored in it, is a rather complicated task too.

This paper presents an approach, intended to create a personalized control model for permitting access to the documents stored in the repository. The basic concepts in it are the strict typification of documents, the creation of role- and operational unit - based levels and the distribution of rights associated with them. This is achieved by way of creating an additional abstract layer containing models of objects and subjects of the repository. Users are granted access rights depending on the role assigned to them, the operational unit they belong to and the affiliation of the document to a certain unit of the organization.

ACKNOWLEDGMENTS

V. REFERENCES

[1] Kasakliev, N., Georgiev, I., Petkov, A. and Gzheferov, Ch. 2007. Electronic University Library, In *Proceedings of the 3rd Scientific Conference*, Technical University of Sofia, Plovdiv, March 2007, pp. 22 - 28.

[2] Stefanov, Kr., Boytchev, P., Stefanova, E., Georgiev, A., Nikolova, N. and Grigorov, A. 2011. Digital Libraries in Teacher Education, Mathematics and Education in Mathematics, In *Proceedings of the Forty Jubilee Spring Conference of the Union of Bulgarian Mathematicians*, Borovetz, April 5–9, 2011, pp. 120-135.

[3] Bothe, K., Budimac, Z., Putnik, Z., etc. 2014. Maintaining quality of software engineering education by a shared repository of course materials in a multilateral setting, In *4th IEEE Global Engineering Education Conference: Engineering Education Towards Openness and Sustainability*, IEEE EDUCON 2014; Istanbul; Turkey; 3 - 5 April 2014, pp. 539-543.

[4] Todorova, M., Hristov, H., Stefanova, E., Nikolova, N. and Kovatcheva, E. 2011. Project-based training on data structures and programming, Mathematics and Education in Mathematics, In *Proceedings of the Forty Conference of the Union of Bulgarian Mathematicians*, Borovetz, April 5–9, 2011, pp. 454-465.

[5] Balabozov, I., Hinov, Kr., Venkov, D. and Yatchev, I. 2017. Automated System for Study of the Dynamic Characteristics of Electromagnets, In *Proceedings of the Technical University of Sofia*, Volume 67, Issue 1, 2017, pp. 381-387, Bulgaria, ISSN 1311-0829.

[6] Luchev, D., Paneva-Marinova, D., Pavlova-Draganova, L. and Pavlov, R. 2013. New digital fashion world, In *Proceedings of the 14th International Conference on Computer Systems and Technologies*, Ruse, Bulgaria – June 28 - 29, 2013, pp. 270-275.

[7] Paneva-Marinova, D., Pavlov, R. and Goynov, M. Two integrated digital libraries for knowledge and iconography of Orthodox saints, In *Lecture Notes in Computer Science(including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7616 LNCS, 2012, pp. 684-691.

[8] Yordzhev, K. and Peneva, I. 2012. Computer Administering of the Psychological Investigations: Set-Relational Representation, *Open Journal of Applied Sciences*, 2012, 2, 110-114. DOI: 10.4236/ojapps.2012.22015.

[9] DocuVantage OnDemand, Document Advantage Corp., http://www.docuvantage.com/document-management-for-higher-education.

[10] Docsvault, Easy Data Access, https://www.docsvault.com/document-management-solution-educational-institutions/.

[11] Treeno, Treeno Software, https://treenosoftware.com/solutions/education/higher-education/.

[12] DynaFile, Blue Ribbon Technologies, https://www.dynafile.com/document-management-software-solutions/education.

[13] eFileCabinet, eFileCabinet Inc., https://www.efilecabinet.com/education-document-management-software/.

[14] FileHold, FileHold Systems Inc., https://www.filehold.com/business-solutions/education.

[15] Virtual Cabinet, Reckon Ltd., https://www.virtualcabinet.com/document-management/industries/education.

[16] Hadzhikolev, E., Hadzhikoleva, S., Orozova, D. 2018. Digital Model of a Document in a University Document Repository, In *Proceedings of the XX-th International Symposium on Electrical Apparatus and Technologies SIELA* 2018, 3 – 6 June 2018, Bourgas, Bulgaria.

[17] InterNational Committee for Information Technology Standards, INCITS 359-2012. 2012. *Information Technology - Role Based Access Control*, http://www.techstreet.com/standards/incits-359-2012?product_id=1837530.

[18] National Institute of Standards and Technology. 2014. *Guide to Attribute-based access control (ABAC) Definition and Considerations*, http://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.sp.800-162.pdf.

[19] The SERES team. 2013. *OrBAC: Organization Based Access Control* (2013), http://orbac.org/.

[20] Zhang, Z., Zhang, X. and Sandhu, R. 2008. Handbook of Research on Social and Organizational Liabilities in Information Security, In *Chapter 6: Towards a Scalable Role and Organization Based Access Control Model with Decentralized Security Administration*, ISBN: 1605661325, pp. 94-117.

[21] Zhang, Zh., Zhang, X. and Sandhu, R. 2006. ROBAC: Scalable Role and Organization Based Access Control Models, In *Proceedings of the International Conference on Collaborative Computing: Networking, Applications and Worksharing*, 17-20 Nov. 2006, Atlanta, GA, USA, DOI: 10.1109/COLCOM.2006.361879.