Issues of Model-Based Distributed Data Processing: Higher Education Resources Evaluation Case Study

Olga Cherednichenko[1], Olga Yangolenko[1], and Iryna Liutenko[1]

[1]National Technical University "Kharkiv Polytechnic Institute", Frunze st. 21, 61002 Kharkiv, Ukraine
{marxx75, olga_ya26, chliv_68}@mail.ru

Abstract. Higher education resources are considered in this paper as a complex heterogeneous hierarchical system. The formal mathematical models for resources evaluation are suggested. The analysis of information and communication technologies applied in higher education establishments is conducted. The necessity of distributed hardware and software infrastructure for data storage and processing in the evaluation activities is shown and substantiated. The distributed data storage and processing architecture is presented.

Keywords. Distributed data processing, evaluation, information system, quality, higher education resources.

Key Terms. ProcessPattern, FormalMethod, Model, SoftwareComponent.

1 Introduction

Nowadays education quality is a crucial factor of both competitiveness of higher education establishments (HEE) and success of HEE graduates in their careers. That's why assessment of education quality is an urgent problem for management activities. The adequate education quality estimates can be the basis for correct management decisions and for realization of improvements. Moreover the estimates of higher education quality may be important for employers and for potential applicants as the main stakeholders of the system of higher education.

Education quality assessment is directly connected with education quality model. Various quality models are used in HEEs around the world [1]. Some of them are oriented on possibilities and results of business processes of HEE. For example, EFQM Excellence Model adopted for higher education introduces two groups of criteria: enablers and results [2]. Enabling criteria cover what the organization does, and the results criteria cover what the organization achieves.

As a rule, HEE plans and manages internal resources in order to support its policy and strategy, and the effective operation of its processes. Resources are the means that provide HEE's functioning. During planning and managing of resources HEE balances its current and future needs.

There are basic and supporting processes in HEE. Educational process is the process which realizes the main function of HEE. It supports university's activities intended for delivering knowledge and skills to students. The resources of educational process include the staff, material and technical facilities and courseware. For example, to teach students some subject, first of all, it is necessary to appoint competent lecturer and a tutor. Secondly, some well-equipped room for lessons should be found. Handbooks, guidelines and other courseware are necessary for representation of teaching material. And finally, to support the educational process such information resources as curricula, announcements and other useful information should be available in a convenient way, for example via the local network or the Internet.

The educational process is justified by academic curricula for different qualifications, such as bachelor and master. The basic unit of academic curriculum is a discipline (fig. 1). Every discipline is described by its syllabus. Syllabus determines all necessary resources for discipline teaching. All HEE's resources are distributed among different units and are allocated on vast territories (fig. 2). Moreover, some resources are in common usage of HEE's units.

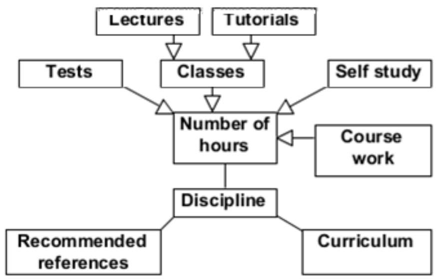

Fig. 1. Discipline structure.

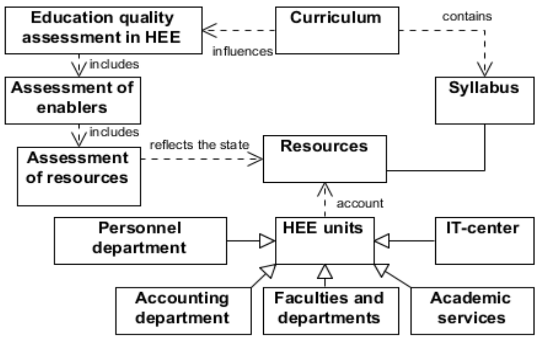

Fig. 2. Resources assessment

So teaching disciplines requires large amount of various resources that are shared, commonly used and accounted by different departments of HEE. Under these conditions assessment of resources that support educational process seems to be a complex problem. The solution of this problem implies collecting data from different sources, its storing, processing and taking decisions about resources updating.

The rest of this paper is organized in the following way. Section 2 classifies information systems in HEE and characterizes mathematical methods used for their realization. Three mathematical patterns that support resources quality assessment are considered in Section 3. Section 4 represents discussion about distributed and centralized data storage and processing. Grid-based architecture is suggested. Section 5 presents conclusions and prospect on future work.

2 Analysis of Information and Communication Technologies in Higher Education

The HEE's functioning is based on various processes, therefore HEE has different information systems (IS) that support these processes. For example, there can be distinguished information systems of administrative and financial management, educational process management and support, scientific research management, information resources management [3]. HEEs use either commercial IS or those which are elaborated by universities for their particular needs [3]. Often commercial solutions do not take into account all peculiarities of definite HEE's functioning, so they can't satisfy all requirements. To elaborate its own software HEE usually has a lack of resources and highly qualified staff. However both types of IS are used and the problem is to integrate them and to provide efficient interaction between solutions of different providers.

The work of all university's IS is based on mathematical models used for solving management problems. Problems of data accounting are managed by means of data bases (DB). Queries to DBs are formed with the help of relational algebra [4]. Data stored in DBs is processed basing on statistical methods which include correlation, variance, discriminant, factor, cluster and other kinds of analyses [5]. Since testing is an important and specific educational problem, processing of testing results is supported by Classical Test Theory [6] and Item Response Theory [7]. HEEs rating and various assessment problems are solved with the help of expert methods [8].

As we can see there are different tasks that should be automated in the university. Different mathematical models and IS are used to solve these tasks. All existing and elaborating IS are intended to the common purpose – to improve university's functioning. From the management point of view all IS should exist in the common informational space with distributed data storage and processing. So we can make the conclusion that it is necessary to develop distributed hardware and software infrastructure composed of heterogeneous resources owned and shared by multiple administrative units which are coordinated to provide transparent, dependable and consistent computing support to a wide range of applications.

Since quality can be considered as one of the main goals of HEE's management, the resources evaluation subsystem has to be a component of HEE's software infrastructure. Resources quality influences higher education quality directly. Our task is to extend the functionality of existing IS by implementing resources evaluation.

Quality monitoring and evaluation (M&E) is a part of the management process. There are two basic types of monitoring: implementation-focused and results-based [9]. Implementation-focused M&E is oriented on inputs, activities and outputs of system's functioning. Results-based M&E focuses on goals and results giving the evidences and explanations of existing tendencies. Our research is oriented on results-based M&E. In the given work we consider the evaluation process which provides management with necessary estimates towards defined management outcomes.

3 Model-Based Resources Evaluation

We consider the system of university's resources as a complex heterogeneous hierarchical structure, which has a large number of parameters. HEE resources are used, controlled and evaluated by different departments and units. We found distributed structure of data sources providing partial estimates of resource units. We distinguish three basic tasks connected to the university's resources M&E. They are resources comprehensive quality assessment, internal licensing audit and resources usage performance evaluation. These tasks are interconnected and related to quality assessment from different points of view.

Higher education licensing, i.e. the process of granting permissions to provide certain educational services, is an important process of public administration. The licensing process is carried out on a regular basis (the license validity is limited), it requires processing of large data volumes, and supposes that the information related to HEE is available for public access. This information in particular may include curricula, syllabus, university infrastructure, etc. The internal audit is required to prove the sufficiency of existing resources and their quality to obtain license. During the internal licensing audit HEE is estimated according to the license conditions [10].

Resources usage performance can be evaluated through the set of indicators – different for each type of resources. For example, performance of capital resources is measured via loading factors, profitability ratio, usage intensity coefficients. The efficiency of human resources usage is determined by labor performance, annual number of workers, economy or excess expenditure of salaries.

The main goal of management in university is improvement of education quality. Resources of the university provide the possibility to reach this goal. Updating university resources is the most realistic way for improvement of education quality. Therefore resources comprehensive quality assessment is the basis of continuous quality improvement. It allows finding the resources that have to be updated, to evaluate their influence on educational process, and to elaborate resources updating projects.

Partial and aggregate indices are used for decision making in all mentioned tasks. Therefore M&E IS must have the model of transition from a set of partial to aggregate estimates. To solve this problem we suggest applying Resources Network System (RNS) [11].

The structure of evaluation system can be presented as an oriented acyclic graph $G_k = (K, A_k)$, where K is a set of nodes, $A_k = (a_{ij})$ determines the directions of arcs, that connect the nodes. The given graph has two types of nodes: node-entries $K^0 = \{k_j \in K : \forall k_i \in K, a_{ij} = 0\}$ and node-aggregates $\tilde{K} = \{k_j \in K : \exists k_i \in K, a_{ij} = 1\}$. Node-entries reflect the partial estimates of resources. The aggregation logic is defined by particular task. Each aggregate is associated with some composite function, or estimator.

In this paper we pay our attention to the task of resources comprehensive quality assessment. We consider direct and reverse tasks of comprehensive assessment. The direct task is the computation of comprehensive assessment value on the assumption of known values of estimators. The reverse task is the determination of values for node-entries on the assumption of user-defined comprehensive assessment.

The ratio scale is used for both types of estimates. We suggest using Qualimetry Theory (QT) for partial estimates computing. The QT provides generalized principles of quantitative assessment of quality of objects of any nature [12]. That's why we choose the method of qualimetry to develop node-entries assessment methodology. To aggregate partial estimates we use the following convolutions as estimators: weighted average arithmetical, weighted average geometric, weighted guaranteed result, and weighted dominating result index. We use different types of estimators because of heterogeneity of resources, different influence of various resources types on the comprehensive resources quality. These convolutions require experts' judgments.

We use the QT-based approach for partial estimates calculation which includes the following steps.

1. Situation assessment is defined, i. e. description of the conditions and goals of the assessment, application of those estimates are identified.
2. The properties tree is constructed. It reflects the hierarchically ordered set of features of the object and allows to fully describe its quality. The procedure of properties tree construction is based on a number of claims detailed in the QT [12].
3. For each simple and some complex features it is necessary to assign appropriate indices. Every index is associated with measurement scale, reference and rejection values. The absolute values of all indices must be converted into relative ones.
4. Weighted coefficients are calculated. We suggest defining weighted coefficients using pairwise comparison. The weighted coefficients calculation is based on the method of eigenvector [13].
5. Partial quality assessment is calculated based on one of the weighted average methods. We use weighted coefficients and relative values of indices obtained on the previous step.

Thus, all resources, that support educational process in HEE, are associated with an oriented graph. Every node of this graph reflects individual or group estimates of resources. We propose pattern-based data processing and distinguish the following patterns.

1. The QT-based Partial Assessment Technique (PAT) is a pattern for assessment of separate resources units. Based on this pattern the evaluation procedure is formed, which results in partial estimate.

2. The Comprehensive Assessment Technique (CAT) defines the grouping of estimates of resources units from different points of view. We suggest to represent CAT as a graph which node-aggregates are associated with one of the convolutions mentioned above.
3. Weighted coefficients Calculation Technique (WCT) provides expert judgment method based on pairwise comparisons. This pattern includes the procedure of processing pairwise comparisons and calculation of weighted coefficients vector.

Resources evaluation is considered on three stages. They are pre-processing, data processing and interpretation of results. These stages are repeated for each evaluation task. The pattern is chosen depending on the task that has to be solved. Initial data sources are determined by RNS. Pre-processing stage includes data extraction from initial sources, collection of required data and its transformation. Data processing stage is totally defined by the chosen pattern. On the results interpretation stage the obtained estimates must be explained from the point of view of the considered task.

Based on the above discussion we suggest creating the resources evaluation software.

4 Architecture of Resources Evaluation IS

Currently the two common ways of data storage are centralized and distributed approaches. Centralized data storage applies that a DB is located at a single server. The data access is executed with the help of remote query or transaction. This is the simplest approach for realization and maintenance. The disadvantage is that the DB would be unavailable for remote clients if connection errors occurred. Also database located on one server has a limited volume of memory. Since all queries are sent to a single server, there are some obvious constraints in time delays and costs of connection support.

Distributed approach supposes that data is stored on multiple servers. This allows increasing of DB volume. Since many queries are satisfied by local DBs, the response time and the costs of query processing are decreased while availability and reliability are increased. The disadvantage of distributed data storage is that some queries and transactions may need the access to all servers which increases the response time and the costs. Another issue is that all clients have to have information about data location in distributed DB. This approach is compatible with common usage of local and global networks.

Distributed data storage and processing provides efficient work with rapidly changing information that is used by various clients. It supports a large number of cooperating clients which collect, register, store and deliver information. Distributed data storage prevents servers from overloading by distributing data among different computers. Also it provides an access to a huge volume of information stored in the system.

One of the most convenient approaches of distributed data storage and processing is grid technology. There are three types of grid systems: computational, data and network [14]. A computational grid has the processing power as the main computing resource shared among its nodes. A data grid has the data storage capacity as its main shared resource. Such grid can be regarded as a massive data storage system built up from portions of a large number of storage devices. Network grid has as its main purpose to provide fault-tolerant and high-performance communication services.

In this work we suggest using data grid architecture for M&E software in HEE (fig. 3). Data stored by various departments and units is shared through the Internet or local network. Data processing is conducted by application servers available for different users via their desktop applications. M&E software services are located on application servers. Special management system is required to control, coordinate and synchronize distributed data storage and processing.

The PAT, CAT and WCT services are based on the corresponding patterns described above. The data collection service provides tools for data extraction from distributed database. These services can be used by quality evaluation software. If we need to evaluate resources quality, this software must store RNS model and data sources for partial estimates. Such software must support pre-processing and interpretation stages. On the data processing stage it calls the necessary services.

We suppose that Open Grid Services Architecture is the suitable solution for distributed hardware and software infrastructure. This will be a base for university's IS integration.

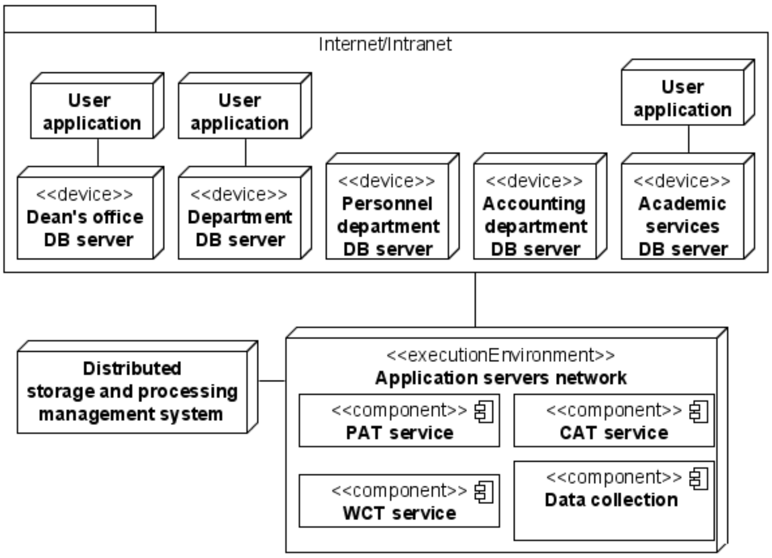

Fig. 3. Distributed data storage and processing architecture.

5 Conclusions and Future Work

We considered a case study of comprehensive resources quality assessment. Higher education resources have distributed and heterogeneous structure. This leads to the idea of distributed data storage and processing. The given paper generalizes methods of quality assessment into several patterns. The suggested approach allows quality assessment of heterogeneous objects and has comprehensive facilities for construction of quality measurement tool. The architecture of information system of resources evaluation is based on distributed data grid architecture. Such approach may be generalized for evaluation of different objects, not only resources and not only in higher education establishment.

Our future work includes elaboration of formal models for solving two remaining tasks – internal licensing audit and resources performance evaluation. The further researches will be devoted to detailed architecture design and software construction.

References

1. Existing models of education establishments' quality systems, http://ru.ict4um.edu.ru/lib/euro/model (in Russian)
2. EFQM Excellence Model. Higher Education Version 2003. Centre for Integral Excellence, Sheffield Hallam University (2003)
3. Krukov, V.V., Shahgeldian, K.I.: Corporate Information Environment of University: Methodology, Models, Solutions. Dalnauka, Vladivostok (2007) (in Russian)
4. Date, C.J.: An Introduction to Database Systems. Addison Wesley (2003)
5. Ross, S.M.: Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press , London (2004)
6. Steyer, R.: Classical (Psychometric) Test Theory, http://www.metheval.uni-jena.de/materialien/ publikationen/ctt.pdf
7. Reeve, B. An Introduction to Modern Measurement Theory, http://www. http://appliedresearch.cancer.gov/areas/cognitive/immt.pdf
8. Meyer, M. A., Booker, J. M.: Eliciting and Analyzing Expert Judgment: A Practical Guide. SIAM, Philadelphia (2001)
9. Kusek, J.Z., Rist, R.C.: Ten Steps to a Results-Based Monitoring and Evaluation System: a Handbook for Development Practitioners. The World Bank, Washington, DC (2004)
10. Cherednichenko, O., Kuklenko, D., Zlatkin, S.: Towards Information Management System for Licensing in Higher Education: An Ontology-Based Approach. In: Mayr, H.C., Karagiannis, D. (eds.) Information Systems Technology and ist Applications 6th International Conference ISTA 2007. LNI, 84, pp. 33--42. GI, Bonn (2007)
11. Cherednichenko, O., Timchenko, K., Liutenko, I.: Technology of Quality Comprehensive Assessment (for Resources in the University by Example). In: Vestnik of Kherson National Technical university, vol. 2 (41), pp. 451--455. KNTU, Kherson (2011) (in Russian)
12. Azgaldov, G.G.: Theory and Practice of Goods Quality Assessment (Basics of Qualimetry). Economics, Moscow (1982) (in Russian)
13. Saaty, T., Vargas, L.: Decision Making with the Analytic Network Process. Economical, Political, Social and Technological Applications with Benefits, Opportunities, Costs and Risks. Springer (2006)
14. Hose, K., Schenkel, R.: Distributed Data Systems: Introduction, http://www.mpi-inf.mpg.de/departments/d5/teaching/ws10_11/dds/slides/DDS-1-print.pdf