ANALYSIS OF LEARNING OBJECT RESEARCH USING FEATURE-BASED MODELS

Renata Burbaite, Vytautas Stuikys

Kaunas University of Technology, Software Engineering Department, Studentu 50, Kaunas, Lithuania, renata.burbaite@gmail.com, vytautas.stuikys@ktu.lt

Abstract. In this paper, we analyze Learning Objects (LOs) from the reusability perspective aiming to better understand the reuse dimension in e-learning and how to handle some reuse issues of LOs more effectively. First, we introduce feature-based modeling concepts borrowed from software engineering for analysis, which outlines dominating factors (features) and the way they affect reusability of LOs. Next, we focus on content/context forming factors such as content granularity, context independence; multiple content/context mappings, accumulative and flexible content updating and changing. Based on knowledge extracted from the analysis, we reconsider feature-based context and content models to analyze and to understand LOs research. We also consider the concept of Generative LOs and motivate the need to extend the concept.

Keywords: reusability, learning object (LO), generative learning object (GLO), context introduction.

1 Introduction

Learning Objects (further LO or LOs) is one of the most important topics in e-learning. A variety of definitions of a LO [1, 2, 3, 4, 5], a variety of taxonomies and standards [6, 7] shows importance of the LOs. Usually a LO is defined as a small, stand-alone, mediated, content “chunk” that can be reused in multiple instructional contexts, serving as a building block to develop high-level compounds (e.g., lessons, modules, courses, etc.) [3].

The idea to constructing e-learning using reusable learning objects (further RLO or RLOs) became a central theme in e-learning research. Reusability is a key property of LOs and can be considered as a strategy to increase productivity, quality, reliability, and to reduce the development costs, etc. [8, 9]. In the field of software engineering (SWE), reuse is treated as the process of creating software systems from existing software rather than building them from scratch (ideal case, box-reuse); in practice, however, reusability requires adding new functionality and modifications (white box-reuse). In the field of LOs, however, there is no single understanding of reusability. There are a wide range of theoretical assumptions on reusability, but real putting into practice is rather slow [8, 9]. We claim that analysis of LOs from the reusability perspective may help to better understand the reuse dimension in e-learning and to solve some reuse issues more effectively.

The aim of this paper is analyze LO research in order to better understand LOs reusability aspects and how reusability aspects can be extended through the concept of generative LOs. The paper’s contribution is: 1) the introduction of a systematic framework for analysis, i.e. feature-based modeling concepts borrowed from software engineering, which outlines dominating factors (features) of the LOs research domain; 2) motivation of the usefulness the feature models for the LO analysis and understanding; 3) illustrative LOs content/context and other feature models.

The structure of the paper is as follows. Section 2 analyses related works. Section 3 presents a description of the feature-based notation. Section 4 motivates the usefulness of feature models for e-learning and LO and GLO research. Section 5 presents case studies with illustrative examples explaining how feature models should be constructed. Section 6 provides a discussion and evaluation. Section 7 gives conclusions and outlines the future work.

2 Related works

Reusability is a key property of LOs. As it is identified in [4], “reusability remains the most difficult to define, since it is related mainly to the instructional design, and not to digital formats or content structure”. This issue is widely discussed in scientific publications [9, 10, 11]. For example, Paris [8], Krämer [9], Sampson and Papanikou [12] define reusability as a strategy to increase productivity, quality, reliability, reduce development costs of LO’s. Silveira et al. [13] note that reuse must increase productivity, maintainability, reliability, extensibility, adaptability. Quinton [14] presents the arguments for designing and developing reusable LOs, which can be identified as aim of reusability that enables the following properties:

- flexibility – material, designed for use in multiple contexts, can be adapted more easily in new context than material, that is rewritten each time to a new context;
- easiness of updates, searches and content management – metadata tags facilitate updating, searching and management of content;
• customization – material can be delivered and recombined at the desired level of functionality and purpose;
• interoperability – the possibility to use LO’s in different learning systems and contexts;
• facilitation of competency-based learning – flexible and adaptive teaching approach with the individual learner’s needs;
• increase of content value – the value of content increases every time, when reused.

The content granularity is an extremely important characteristic of LOs because it enables to aggregate higher-level compounds (e.g., lessons) from smaller parts [4, 15, 16, 17]. Silveira et al. [13] define the relations between the learning source (element) type, granularity and reuse, Balatsoukas et al. [18] offer content models with different levels of aggregation. We summarize all these concepts in Table 1.

Table 1. Relations between element type, granularity, aggregation level and reusability

<table>
<thead>
<tr>
<th>Learning content type</th>
<th>Granularity level</th>
<th>Aggregation level</th>
<th>Reusability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw media: images, audio-video files, text snippets</td>
<td>very fine</td>
<td>low</td>
<td>Reused on the “use-as-is” basis</td>
</tr>
<tr>
<td>Anything that focuses on a single piece of information</td>
<td>fine</td>
<td>low</td>
<td>Reused as self-contained units. Can be disassembled in order to reuse their content assets</td>
</tr>
<tr>
<td>Collection of information objects that are assembled to teach a single learning objective</td>
<td>medium</td>
<td>medium</td>
<td>Reused as self-contained units. Can be disassembled in order to reuse their information objects or content assets</td>
</tr>
<tr>
<td>Lessons, courses, general learning resources composed by multiple LOs with multiple learning objectives</td>
<td>coarse</td>
<td>medium</td>
<td>Entirely reused, but coarse granularity reduces reuse potential</td>
</tr>
<tr>
<td>Learning environment – combination of content and technology with which a learner interacts</td>
<td>very coarse</td>
<td>high</td>
<td>Reused depending on the coupling they maintain among them</td>
</tr>
</tbody>
</table>

Context is another important characteristic influential to reuse. Usually context is defined as a complex concept, which may include a range of factors such as composition and complexity of learning materials, types of social situations, constraints of learning tools, types of learning activities etc. [19]. Context models are necessary in order to create effective learning: “the more context a learning object has, the more (and the more easily) a learner can learn from it” [20]. On the other hand, for reuse to be successful, the LO should be independent of context [4, 15]. Damaševičius and Tankelevičienė [19] propose Learning Context Model, which is aiming „to represent variability of learning context, and to support flexibility, reusability, adaptivity and personalisation“. Man and Jin [21] discuss the context rich paradigm. The concept “context rich” means that we should “include sufficient context information into the description of the learning object and various services, especially the dynamic information always preserved by the services”. The authors define the context model, which consists of three parts: context information from the user, context information from the service and the content information. Huddlestone and Pike [22] point some context limitations to support reusability (e.g., equipment, procedural, legal, doctrinal, cultural or national, environmental, climatic or situational). In order to enforce reusability, Quinton [14] defines LO contextualisation strategies as follow. It is possible to generate:

• multiple context wrappers in a learning environment;
• context frames, corresponding to the needs of learner – tailored context frames; adding context links to object;
• pattern templates, which are used to contextualize LO in relation to the abilities, knowledge in performance-based approach to using LOs; e.g., Jones and Boyle [23] propose to use LO design patterns with purpose to improve reuse the LO or adapt it for use in a different context.

The following papers analyze the dependency between content and context. A flexible content is more adaptable in different contexts [15] and can be used in various learning activities and environments [14]. The pedagogical model is a part of pedagogical context [6, 15]. LOs, which are independent of the pedagogical model, have a greater potential of reuse.

Social factors are also important to consider. For example, motivation is an essential factor for success or failure. Cocea and Weibelzahl [24] identify factors that have the greatest impact on motivation. Some of these factors have an impact on reuse, e.g. feedback based on performance, personalization, adaptivity, collaborative
learning. Next, organization culture is seen from the perspective of reuse as an intention to share the content and the work collaboratively [22]. Learning scenarios are understood as a part of social context [25]. Pernin and Lejeune [26] propose the learning scenario sets that improve reusability in different contexts. From the perspective of reuse, educational theories are seen as a part of educational (pedagogical) context. A variety of LOs types would support the use in a variety of educational theories. Copyright must be considered not only from the aspect of using existing material, but also from the view of the creator’s own perspective [27]. Open educational resources can be successfully reused too, but it is necessary to create a critical mass of the content that is needed to support reusability [28].

Štuikys and Damaševičius [17] propose to develop GLO using feature diagrams (FDs) and generative techniques. Authors describe design principles to design GLOs as a sequence of activities: 1) analysis of LOs and its context; 2) representation of the domain by FDs; 3) introduction of a relevant generative technology; 4) selection of languages for implementing, 5) development of the model of GLOs, 6) transformation of high-level model and GLO model by merging them into a coherent model using some transformation rules; 7) specification contains meta-interface (serves to specify metadata extracted from FDs model) and meta-body; 8) verification of the specification, 9) transferring the verified specification into web-based environment (server), 10) on-line verification of GLOs through organizing sessions of links “server-PC” and testing of how the generated LO instances correspond to the metadata values and user requirements (without feedback).

Analysis of related works allows us to summarize the following. There are two general views to consider reusability of learning objects. The first is defined as a local and the second is seen more or less as systematic. The first view focuses on the component-based view and concentrates on a few reusability aspects (e.g., granularity, compositionality, etc.). The second view expands the number of aspects. Furthermore, this view combines the LOs structural aspects with the LOs design aspects. However, there is a lack of a well-grounded systematic approach to deal with reusability of GLOs. As a result, in Section 3, we introduce a more comprehensive model as a framework for systematic analysis and reuse understanding.

### 3 Analysis of feature-based notation

We have borrowed the feature-based notation, i.e. Feature Diagrams (FDs) from SWE [29]. We found FDs as a very useful notation to use it in e-learning due to many reasons which we will explain later. Now we discuss the notation itself.

#### Table 2. Feature types for feature model representation

<table>
<thead>
<tr>
<th>Feature type</th>
<th>Definition, formalism and semantics of relationships</th>
<th>Graphical notation (syntax)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mandatory (all relationship)</td>
<td>Feature B (C, D) is included if its parent A is included a) if A then B; b) if A then C&amp;D; Father A has son B; or Father A has sons C and D</td>
<td><img src="image1" alt="Diagram" /></td>
</tr>
<tr>
<td>2. Alternative: one-of (aka or-selection)</td>
<td>Exactly one feature (B or C or D) has to be selected if its parent A is selected: a) if A then case of (B, C); b) if A then case of (B, C, D) Only one son from the list (B, C) or from the list (B,C,D) is selected; Relationship “case of”</td>
<td><img src="image2" alt="Diagram" /></td>
</tr>
<tr>
<td>3. Alternative: sub-type xor-selection</td>
<td>if A then (B but ¬C) or (C but ¬B) Differs from case -selection by 1) having two sons only 2) label “xor” usually written below the father’s node</td>
<td><img src="image3" alt="Diagram" /></td>
</tr>
<tr>
<td>4. Constraint exclusive-or</td>
<td>if F then ¬K and if ¬F then K (here F and K are atomic features derived from different parents); Mutual exclusive features (xor relationship between F and K) Line between boxes marked by tag xor for features derived from different parents</td>
<td><img src="image4" alt="Diagram" /></td>
</tr>
<tr>
<td>5. Constraint require</td>
<td>Feature A requires feature B, or shortly: A (require) B Line between boxes marked by tag require</td>
<td><img src="image5" alt="Diagram" /></td>
</tr>
<tr>
<td>6. Constraint equivalence</td>
<td>A require B &amp; B require A ⇔ equivalence(=)</td>
<td><img src="image6" alt="Diagram" /></td>
</tr>
</tbody>
</table>

Informally, a feature is a prominent characteristic of a system, entity or concept in a domain. Since in the software engineering literature there is no consensus on what a feature is, we deliver some definitions of the term. A feature is: (1) End-user visible characteristic of a system or a distinguishable characteristic of a concept that is relevant to some stakeholder [30]; (2) A logic unit of behaviour that is specified by a set of functional and quality requirements [31]; (3) Qualitative property of a concept [32]; (4) A functional requirement, or a reusable product line (PL) requirement, or a characteristic that is provided by one or more members of a software PL [33].
All these notions of features are valid and more relevant for different PL methodologies that are dealt with within Software PL engineering which is “a paradigm to develop software applications (software-intensive systems and software products) using platforms and mass customization” [34].

A feature-based model is the one that consists of two kinds of elements (features and relations), where features are represented by nodes (i.e., boxes with marks meaning the kind of features) within a feature diagram and relations are represented either by branches (parent-child relationships) or by constraints of type REQUIRE and XOR, meaning constraining relations among feature children having different parents (see Table 2).

The root represents the top level feature, i.e. the whole LO. Intermediate nodes represent compound features and leaves represent atomic features, also called variants that are non-decomposable. There are three types of features: mandatory, optional /alternative. Mandatory features are denoted as boxes with black circles, alternative (optional) – as boxes with white circles (see Table 2). Mandatory features allow expressing the commonality of the concept, optional and alternative features allow expressing variability (those terms will be explained later). Features can be a solitary or be represented in groups using arcs on branches (see Table 2, 2-nd row).

4 A motivation to use FDs in e-learning

As there is no wide acceptance of the notation in e-learning so far, we need to motivate the role of FDs more extensively. From the perspective of teacher and course designer, FDs are important due to the following reasons:

- Capability to specify LO at different abstraction levels.
- Adaptation for different needs (teaching contexts, content changes and extensions).
- Easiness for communication (FDs can be seen as a specification document to be transferred).

From the perspective of the topic, such as Computer Science (CSc)-related topics, FDs are important due to:

- FDs are build on the well-grounded theoretical foundation (i.e., on propositional logic and set-based and graph theories), thus the notation is theoretically sound.
- Each of those themes are also learning topics (i.e., LOs) within CSc courses, thus the use of FDs by instructors in teaching has yet another explanation and motivation.
- The FD per se can be seen as a LO within the courses, thus the use of the same notation for representing LOs and teaching, in our view, empower the motivation of learner to learn.
- CSc-based courses has many formal knowledge and this knowledge should be expressed at various formalization levels, for example, from meta-knowledge (i.e., very general) to very concrete; thus FDs can be seen as a well-suited tool for this purpose.
- FDs have three representation forms: set theory based, graphical notation (FD), textual representation as a Feature Model Language (FML); thus it can be adapted to different education levels (e.g., school, colleagues, University).

From the perspective of the generative learning objects (GLOs), FDs are important due to the following reasons:

- Explicit representation of variability within both content and context because of variability is a core property to describe and understand the GLO concept.
- To manage variability changes and extensions.
- To manage complexity of GLO.
- FD is a specification document to understand the concept and to realize it.
- FD specification supports evolution of GLOs within their life cycle.
- Feature-based specification of GLO by no doubt is a research topic.

The provided discussion also evaluated the advantages of the notation. The full evaluation requires to state disadvantages of the notation too. They are as follows:

- Less maturity in comparison with other similar notations (e.g., UML).
- Lack of adequate tools to support drawing and interpretation.
- Some semantics discrepancies within various with graphical notations for edge, branch coloring.
- Some difficulties are to make changes, to read large documents given on different pages.
- Preparation of feature-based LOs with computer is a time consuming process as it requires a careful attention and concentration from the designer.
In our context, feature is a knowledge unit (concept) used to understand LO research through modelling (i.e. constructing and analyzing high-level feature models). Feature-based model is the model which describes a domain (some accepted aspects of LOs with the domain scope in mind) as generally as possible using the feature-based notation. Instantiation of the feature model is a form of model transformation that includes: 1) selection of features from its model and 2) decomposing of the selected features into sub-features until variant features are derived. Variant feature is the feature value beneficial for use or understanding. Partial instantiation is the transformation in which the only part of features from its model is selected and variant features are derived from the selected features. As it was identified in Section 2, at the core of learning are the context and content. In Section 5, we present their models.

5 Case studies

The aim of case studies is to motivate some statements given Section 4 and show with examples how feature-based models can be constructed and used at the different abstraction levels (from high to low). First we present the high-level context model separately from its content model because of the following reasons: 1) not all context features are important for a concrete learning content (e.g., GLOs in our case); 2) when context features are selected (to satisfying some learning objectives), it is possible to identify relations between the context and the content and combined both models. Then we present the content model combined with the selected context features. Both models (Case 1 and Case 2) are treated as high-level illustrative models enabling to provide the instantiation process in order to construct lower-level models with concrete context and content (for example, for the specification of GLO “Sorting” before its development). Finally we present the feature-based specification of GLO “Sorting” (Case 3).

5.1 Case 1: LO domain context model

Whatever model we have, the context model or content model, the first task is to identify the scope of a model. By scope we mean the number of features in horizontal dimension and the number of features and sub-features in vertical dimension (see Figure 1). That depends on analysis objectives and tasks. As in our case objectives are very general, i.e., understanding of LO research trends and the concept of GLO, the context is wide enough but it is by no means exhaustive. For example, we have introduced the following context features, which are outline in Figure 1. The model is constructed using: 1) explicit or implicit knowledge gained from related work studies as they are presented in Section 2 (of course, partially) and 2) formalism given in Table 2.

![Figure 1. Feature-based context model to understand LO research](image)

In our view, pedagogy theories, e-learning standards, e-learning processes, LO repositories, actors, their competence, LO reusability, technology support are high-level features of the general context model. What is important for the model understanding is: 1) relations among context features within the context model (note that some Require relations are missed in Figure 1 due to simplicity reasons); and 2) the identification the features which require content features (e.g., the following context features: <LO design>, <LO reusability>, <Technology support> require content features aiming to construct a GLO specification).

5.2 Case 2: LO content model

The model is given in Figure 2. The model represents the only those features that are important to understand the concept GLO and its relations with context model and LO reusability in general. We have already identified the context model features that are influential to understand GLO. Here we describe them using the require relation more precisely as follows:

- <LO design> require <content>
• <LO reusability> require <content>
• <Technology support> require <content>

As there are some problems to represent relations among leaves within tree-based feature models, the only part of those relations are depicted in Figure 2. To understand the content model we need to define some concepts given within the model.

LO commonality is a set of features or a unique feature that are fundamental piece of knowledge, i.e., common in different contexts of use (e.g., three assignment statements to perform the permutation of two items in different sorting algorithms \((x = a_i; a_i = a_j; a_j = x)\)).

LO variability is a set of feature variants that differ in different contexts of use (e.g., sorting order (increasing, decreasing), algorithm type (bubble, insert, etc.), programming language type (Pascal, C, C++, etc.)).

GLO is a set of the related LO instances with different features (i.e., having high variability) realized using some generative technology (e.g., template-based, meta-programming-based, aspect-programming-based, etc.).

Figure 2. LOs' content model (BB- black box, WB- white box, GB-glass box, GLO - generative LO)

5.3 Case 3: Feature-based model to represent knowledge fragments

This case study demonstrates how knowledge about sorting can be represented using a feature model. The model contains the following features: List and Operation (see Figure 3). The latter contains the following sub-features: Selection of two elements of the list and comparison (this feature is specified as "Selection & Comparison" shortly in Figure 3), Permutation, Next pair, End of selection.

Note that: 1) this model is described with some redundancy because it is illustrative example and we are aiming to show expressive power of FDs; 2) the model is not intended for teaching or learning (for this purpose it must be simplified); 3) in order to understand the essence of the feature “End of selection” some knowledge should be introduced from the context (e.g., previous knowledge about dependent loops of the type: “for… for … end for… end for”).

Figure 3. Essential knowledge represented using FD to understand sorting (a) and variability aspects (b)

5.4 Case 4: Variability model for specifying GLO “sorting algorithms”

This case study is given in Figure 3, b. We hope that it is understandable without extended explanations. For simplicity reasons, Require relations are not shown in Figure 3, b.
6 Discussion and evaluation

Though it is commonly agreed in the e-learning community that reusability is one of the most important attributes of LOs, so far, however, there is no unique view on the reuse issues, firstly, because of the topic wideness and, secondly, due to the lack of a well-grounded approach to deal with the topic. In our view, in order to better understand reusability of LOs, it is not enough to extract characteristics affecting reuse; it is more important to try to express characteristics as accurate as possible and represent the relations among the characteristics explicitly. We borrowed the feature-based notation from Software Engineering and found the notation beneficial for representing LOs per se and their reuse characteristics within the proposed feature models explicitly. Though the models, by no means, are comprehensive, however, they outline the most essential reuse dimensions in the whole. Furthermore, the models provide the possibility for a more detailed classification.

Our specific interest was to analyse and better understand the reuse issues of generative LOs (GLOs). From the technology-based view, the concept GLO means generative reuse, when LO instances are generated from the GLO specification automatically on demand. The primary intention of introducing of GLOs was to enhance reusability through the content adaptations and the capabilities to do that automatically (or at least semi-automatically). Analysis convinced us that GLOs extend the concept of reusable LOs essentially. In general, the design and the use of LOs highly depend on the context, which is extremely wide in e-learning. Our analysis showed that the importance of GLOs for e-learning is already recognized (see, e.g., papers Boyle et al. [35], Morales et al. [36]), Štuikys and Damaševičius [37], Han and Krämer [16]), though the extent and effort in GLOs research is much less in comparing to the so-called reusable LOs, or component-based LOs or simply LO instances.

We have also identified through analysis that the concept of GLOs has the possibility to introduce and to manage the context-based features at the much larger degree. Even more, the preliminary analysis of context and content features has leaded us to the conclusion that the features can be combined together into a coherent structure, if a suitable technology is applied to realize the GLOs. As a result of the provided analysis, we have identified the need for the further extension of the GLOs concept and have formulated the research topic as extended GLOs. Feature-based models can be understood intuitively; they don’t depend upon a domain and allow expressing and managing content and context variability explicitly. However, the models have less maturity and some difficulties in representing and interpreting automatically.

7 Conclusions

We have analyzed the reuse issues in a variety of dimensions but mainly we were focusing on representation aspects of LOs. Our main contribution is an extended motivation of the usefulness of FDs notation in e-learning in general and for the specification, knowledge representation within LO and design specification in creating the generative learning objects in particular. The proposed feature-based framework to represent reusability issues explicitly enables to analyse and understand the issues systematically. As feature models enable representing learning context and contents variations explicitly, they can be considered as a suitable instrument for specifying GLO at a higher abstraction level.

References


- 208 -
THE USE OF INFORMATION TECHNOLOGY TOOLS FOR IMPROVING THE QUALITY OF LEARNING

Aleksandras Targamadze¹, Ruta Petrauskiene²

¹Kaunas University of Technology, Faculty of Informatics, Studentu Str. 50, Kaunas, Lithuania, aleksandras.targamadze@ktu.lt
²Alytus College, Distance Study Centre, Ligonines Str. 1c, Alytus, Lithuania, ruta.petrask@gmail.com

Abstract. Information technologies (IT) tools have been frequently applied in the learning process to improve its quality, but not always this process gives optimal results and it is not always clear for the teachers who select these tools. Although manufacturers offer various IT tools which are characterized by a wide range of possibilities, usually they are not used according to the reasonable method or demand but according to habitual experience of the teachers or as an experiment. A short review of IT tools is presented in the article as well as the experimental research the purpose of which was to analyse the selection of IT tools for a course.

Keywords: distance learning, virtual learning, information technologies in learning.

1 Introduction

Modern educational environment is closely related to ITs. Many scientists have researched the application of ITs in learning ([10], [4], [16], [6], [3] and other). The objects of their research were educational environment based on IT, the application of IT, possibilities of using IT for education, connections with modern training methods, etc.

Historically every new technology stimulated new expectations and blazed the trail for the further technologies. Books, computers and computer networks, the internet – every technology in its time was the most advanced and the most acceptable for the learners and evoked revolutionary changes in learning.

Currently virtual learning environments (VLEs) are well-established in learning. They integrate tools for supporting study process, educational material tailored to students, opportunities for knowledge testing, communication tools. World Wide Web supports the information browsing and the systems of e-libraries with full-text document databases; the internet gives the opportunity of connecting to VLE where one can study material, work, test knowledge, take examinations. Video conferences provide suitable conditions for transmitting video and sound, and ensuring communication relations between students or between the teacher and students. Usually several tools are used in educational process, for example, not only video conferences and VLE, but also e-mail, discussions, etc. Technologies are estimated according to how they manage to convey the content for the development of knowledge and skills as well as the support of interaction processes.

But in practice teachers usually select IT tools for the course (study subject) only according to their own experience and computer skills but they do not estimate educational suitability of these tools. The methodology of selecting IT tools for the course is unknown for the majority of the teachers and it is not used while the analysis of possible application of IT tools suitability has not been performed either. Sometimes that does not improve the quality of studies but on the contrary – that might worsen it.

[15] emphasizes that teachers who create learning content often do not have essential competencies for selecting technological solutions to implement the content. According to the authors teachers must use the recommendations of the technology professionals. On the other side, according to [12], none of the IT tools can support all learning methods, the most effective way is to use various technologies and combine them. It follows that the selection of IT tools demands precision, knowledge and learning process analysis.

In this article an experimental research is presented, the goal of which is to analyse the selection of IT tools for a course using the method of the IT tools selection. The method consists of 4 stages. In the first stage the course goals, learning conditions and quality requirements are evaluated. The learning method is selected according to the evaluation. In the second stage educational activity types are selected and a set of educational activity types is determined for every course goal according to the learning events describing the learning process, with reference to Leclercq and Poumay [11]. In the third stage IT tools are selected for every activity type to enable its implementation. And in the fourth stage a set of IT tools for the course is composed.

The teachers of Alytus College who were involved in the EU structural funds project together with Kaunas University of Technology participated in the research and selected IT tools for their improved or newly designed courses. Using questionnaire the teachers evaluated the peculiarities and benefits of this method, as well as presented information about their experience in applying IT tools. The selection of IT tools was analysed on the course level.
2 IT tools in learning

First distance learning studies used mail technologies, printed or hand-written texts were used to be sent. Desire to improve and increase efficiency of learning has changed through communication tools, and later – trying to secure feedback – using the internet, e-mail, File Transfer Protocol (FTP) server service for data storage and usage, virtual learning environments, etc.

At present ITs is inseparable part of learning. The following tendencies of IT application can be distinguished:

- Enrichment of the learning content with simulation processes through creation of the conditions for transferring of interaction with real practical equipment to virtual space.
- Increase of learning accessibility and reach (t-learning, m-learning, u-learning, etc.).
- Facilitation of information search and usage of learning material, distance learning material creation, learning material cost cutting (semantic webs, learning objects).
- Intellectualization of learning process, transmitting human relation functions which are essential in the learning process to the objects.

Nevertheless it is noticed that there is no united system or methodology how to implement learning process through ITs. Most researches are usually intended for searching new applications of ITs in learning and for experiments. In practice IT tools known and usable by teachers, without evaluation of the conditions and learning goals are usually used. Therefore IT tools are often used unreasonably and because of the big variety of potential IT tools educators are uncertain which tools in what situations are advisable to use.

At the moment blended learning dominates using the combination of synchronous and asynchronous learning methods. Blended learning method is usually implemented by using various virtual learning environments and course managing systems (Blackboard Vista, Moodle, Authoring Tool, LearningSpace and other).

Simulation processes are used to develop and improve practical skills. Simulation programmes replace practical usage of laboratory facilities. Simulation often replaces real or artificial situation which is difficult to create or control in the laboratory. Real laboratory process is transferred to virtual learning environment in the internet and helps to understand the mechanisms which control the simulated system, creates possibilities to recognize the process herewith to increase students’ motivation [2].

Practical processes of simulation were analyzed by [7], [5], [8] and other scientists. Usage of simulation is convenient because students while using software and changing parameters can investigate the processes, analyse equipment reaction, experiment a lot, get instant experiment results without physical and financial resources. Video cameras can also be used and real laboratory equipment can be controlled by remote mode. Use of video recordings allows students to independently analyse the process, clearly see the result changes or work mistakes. Although simulative work has its limits as compared to real objects, virtual laboratories are widely used in learning. The internet connection with the real equipment allows performing remote laboratory works using traditional laboratory equipment.

Also intellectual technologies are inserted into the training practical processes of learning: software agents can coordinate system, choose optimal object management scripts, help students to master the knowledge. Such system is presented by [1].

With the growing amount of the information on the internet the problems of information excess, complicated information management and search and information analysis occur. This problem can be solved with the help of semantic web. In the semantic web information is stored according to the meaning; semantic web “understands” the request according to the notional content connections and the connections with other external objects and solves it more precisely herewith facilitating and specifying the search.

Learning objects theory and practice are more and more expanding in learning. Learning objects are digital units which can be repeatedly used in different contexts. Learning objects allow flexible creation of learning courses and provision of convenient learning opportunities; they can be reached anytime from any place, where there is the internet access [13].

The connection and supporting tools are getting cheaper and mobile and remote technologies are more and more used in learning: t-learning (tele-learning), m-learning, u-learning (ubiquitous learning).

Another area of technology application for learning is usage of software educational agents in virtual learning environments. Software agents can intellectualize learning, individualize it, and support social context and interaction.

[9] states that influence of technologies upon teaching and learning depends on a lot of elements and connections between them. [14] emphasizes that selection method of media or communication for learning provision must refer to systematic approach to distance learning emphasizing the priority for multimedia, which usage is based on analyses of students, content and projection.
It is important to realize that usage of technologies in learning must be purposeful because IT usage itself does not add more quality to learning process. IT “does not automatically add quality to teaching and learning”, – states [6]. Sometimes IT opportunities to reach huge information amount can disbalance learning process if there are no definite and clear methods and tools. It must be clearly understood which IT tool must be used that knowledge transaction wouldn’t be within the frame, which suppress creativity. Therefore IT application must be based upon principles of didactics.

A lot of patricians emphasize communication problems in contemporary learning. Activation, grouping and encouragement of students are often problematic in traditional studies as well. Virtual environments in use moves social interaction to another area where activity must lean upon personal interest and/or reliable teacher-student(s) interaction: activation, involving into discussions or problem solving, reminding, directing to read valuable or interesting material, presentation of information about other students, etc. But this field in the context of IT use is still problematic.

Although it is obvious that IT usage in learning has both – advantages and threats, the majority of the scientists, who research the interaction of IT and learning environment, emphasize inevitability of modern technologies. On the other hand, an effective application of ITs is still an open and insufficiently researched question because it is based on practice more than on scientifically based methods, analysis of the learning process and rational selection of IT tools. It is noticed that not all IT tools allow individualisation of learning, provision of support and materialisation of essential social interaction in learning process. Therefore in learning it is attempted to apply artificial intelligence based technologies which would react to the learning process events and do it adequately to situation.

3 Research on selecting IT tools for a course

The purpose of the research was to evaluate the situation of how teachers select IT tools for their courses. 23 teachers of Alytus College who worked on the EU structural funds project “Increasing quality and internationality of First cycle study programmes in the Economics and Management Faculty of Kaunas University of Technology and in the Faculty of Management of Alytus College” in July 2010 participated in this research, applied IT tools for their courses and expressed their opinion.

During the experimental research the method of selecting IT tools for the course was introduced to the teachers based on which they selected IT tools for their courses. A questionnaire was used to learn teachers’ opinions about the selection of IT tools. Experimental research was carried out and the overview presented in three aspects: 1) Teachers experience in selecting IT tools for their courses; 2) evaluation of the method of selecting IT tools for the course; 3) attitudes towards selecting IT tools in the future.

In the first group of the questions related to the teachers’ experience in selecting IT tools for a course the respondents were asked to answer the following questions: “Have you ever selected IT tools for a course before?”, “How did you select IT tools for a course?”, “Is there a lack of methodical and technical knowledge while selecting IT tools?”

The research revealed that the question of selecting IT tools is a topical issue for teachers because the bigger part of them must select IT tools for their course (61% of the respondents selected tools for their courses, 39% of the respondents – party selected). The survey showed that the implicit situation was confirmed: the IT tools that are usually used are selected according to only previous experience (83% of the answers) or after consulting with colleagues (52%), and sometimes – by experiment (17%). The answer that IT tools were selected according to methodical recommendations was indicated only by 13% of the respondents (3 persons of 23) (Figure 1).

![Figure 1. Distribution of the answers to the question “How did you previously select IT tools for your course?”](image-url)
“partly” – 35% of the respondents) (Figure 2). The assumption can be made that though the selection of IT tools in contemporary studies is a topical issue and it constantly comes into question, the bigger part of teachers do not know the methods of selecting IT tools or do not apply them and feel the lack of such knowledge as well.

In the second group of the questions, devoted for evaluation of selecting IT tools for a course, the following questions were given: "Do educational activity types by Leclercq and Poumay fully suit to implement learning process activities for a course?, “Evaluate the stages of selecting IT tools” and “What IT tools did you select for your course?”. The educational activity types by Leclercq and Poumay who propose learning process as a set of eight types of activities (events) happening in learning, and the stages of selecting IT tools were described in the methodology of selecting IT tools and presented to teachers.

In the opinion of the bigger part of the respondents educational activity types by Leclercq and Poumay fully suit to implement learning process activities for a course – 70% of the respondents answered “yes”, “partly” was indicated by 26% of the respondents.

The respondents’ opinion about selecting IT tools for a course while separately evaluating all four stages of the method of selecting IT tools for a course was also positive: bigger part of the respondents considered such selection of IT tools as clear and rather easily applied (average evaluation of all four stages was 38%). The best evaluated stage of all four IT tools application stages was the fourth one – the formation of IT tools set for a course (74% answered that it is clear and easily applied), Figure 3.
desktops (7 answers) and students’ e-portfolio (4 answers), etc. The selected IT tools according to the presented examples showed that although some of the teachers attempt to involve more various tools which develop technological and communicative student’ skills into learning process, a large variety of tools is not fully used.

While communicating with the teachers orally they made the remarks that some of the IT tools were absolutely incomprehensive and what was more unclear – how to usefully incorporate such tools into learning process. On the other hand some teachers indicated that some IT tools which were created not for studying purposes but for broader circle of users can be successfully involved into learning, e.g. various calculation programmes downloadable from the internet – VAT calculation, wage, pollution tax, bank interest calculators and other calculators.

In the third research tool question group which was intended for learning about the attitudes of the respondents on selecting IT tools for a course the following questions were given: “Was the method of selecting IT tools for a course useful for you?”, “Are you planning to refer to it in the future?” and the open question which asked to propose remarks and comments about selecting IT tools.

The answers revealed that the method of selecting IT tools for a course is evaluated as useful (“yes” was given by 91% of the respondents) or partly useful (9%). 78% of the respondents plan to use the applied method of selecting IT tools in the future. Other evaluations were also positive. In the provided comments it was emphasised that the selection of IT tools is still a topical issue and it was not clear how to select the tools and the methodology was missed. The opinion was given that the application of the method of selecting IT tools is a tool for improving the study process which forces reconsidering the teaching methods, purposefully selecting IT tools and which allows improving the quality of learning.

4 Conclusions

1. ITs are inseparable part of modern learning. The following tendencies of IT application can be distinguished: enrichment of the learning content with simulation processes through providing the conditions for transferring of interaction with real practical equipment to virtual space; increase of learning accessibility and reach (t-learning, m-learning, u-learning, etc.); facilitation of information search and usage of learning material, distance learning material creation, learning material cost cutting (semantic webs, learning objects); intellectualization of learning process, transmitting human relation functions, which are essential in the learning process to the objects.

2. With the growth of the variety of IT tools a part of teachers faces the problem of which IT tools to select, apply and why. It is noticed that there is no united system or method how to implement the learning process through ITs. Therefore in practice IT tools known and usable by teachers without evaluation of the conditions and learning goals are usually used. It is not evaluated how to achieve the quality of learning.

3. The proposed method of selecting IT tools for a course which allows easier and more purposeful selection of IT tools for a course was used in Alytus College by the teachers selecting IT tools to improve or create their new courses during the project.

4. The teacher survey revealed that the method of selecting IT tools for a course motivates them to go deep into the learning process, to understand its parts, the activities and interaction between the student and the teacher. The analysis of IT tools facilitates evaluating the functionality and characteristics of the tools and perceiving the potential and purpose of their application.

References


Abstract. The assessment of knowledge and competences is an integral part of the learning process. It summarizes the level of student success in achieving the corresponding outcomes and objectives. This paper focuses on analyzing electronic exam data using clustering and association rules in data mining. The aim of this work is to obtain and interpret the results and propose solutions to improve the e-examination system using a descriptive model.

Keywords: data mining, e-examination, clustering, association rules.

1 Introduction

Currently, Lithuania's higher education system is reformed, and the basic idea of the reform is to create a free market of higher education, which would create a natural competition between higher education providers. Nowadays higher education providers might be state universities, independent universities, state colleges and independent colleges. The essence of the reform is the student's basket (committed money for a student's studies paid by the state to a higher education institution, where the student is enrolled). The promotion of this basket encouraged higher education institutions to compete for students, who would study in a chosen higher academy and thereby would ensure the sponsorship of this high school. The resulting competition led the higher education community to seriously look into these aspects of the study process, which most affects students’ satisfaction with the learning process.

During discussions with students about the quality of the courses, students, defining their expectations and desires, often point out that they want and expect: interesting and professionally useful lessons, an objective, fair and transparent assessment, modern laboratories for acquiring practical skills, study materials and additional literature available for all.

The purpose of this article is to choose right data mining techniques for analysis of students' e-examining data in order to explore students’ behavioral characteristics whereas having the exam electronic way and in accordance with the results to offer recommendations for a higher quality of exam arrangement and organization of examination.

Research methods: the analysis of scientific literature, data analysis using data mining techniques.

2 Literature review

Data mining means searching for certain patterns within large sets of data, which creates a lot of possibilities for decision makers. By analyzing those patterns, better decisions can be made in order to improve learning and assessment process. The research interest in using data mining in e-learning is constantly increasing. According to L. Shen, M. Wang, R. Shen the database of learning management system includes much useful information which can be effectively used for the improvement of e-learning process [11]. E.Garcia, C. Romero, S. Ventura, T. Calders emphasise that learning management systems accumulate a vast amount of information which is very valuable for analyzing the students’ behavior and could create a gold mine of educational data [4]. Authors emphasise that due to the vast quantities of data these systems can generate daily, it is very difficult to analyze this data manually and a very promising approach towards this analysis objective is the use of data mining techniques [4].

Using data mining methods many kinds of knowledge can be discovered [5]. The discovered knowledge can be used to better understand students' behaviour, to access student’s learning style [7], to adapt a course content according to student’s knowledge and abilities [6], to assist instructors, to improve learning and teaching process [5], [1]. Literature describes a number of scientific research works which use data mining methods on e-learning data. A. Merceron, K. Yacef present a case study that uses data mining methods to identify behaviour of failing students to warn students at risk before final exam [8]. V. Namdeo, A. Singh, D. Singh, R. C. Jain compare different classification algorithms and check which algorithm is optimal for classifying students’ based on their final grade [10]. Other authors use neural networks for predicting student’s marks [2].

Different data mining methods are used for e-learning data analyze, the most common ones are: association, classification, clustering and outlier detections [5]. Literature describes various algorithms of these methods. The choice of data mining method and its realizable algorithm depends on available data, set research goals and intended results [10]. According to A. Merceron, K. Yacef association rules are very useful in
educational data mining since they extract associations between educational items and present the results in an
intuitive form to the teachers [9]. Authors emphasise that association rules require less extensive expertise in
data mining than other methods [9].

3 E-examination system of Vilnius Gediminas Technical University

The assessment of knowledge and competences is an integral part of the learning process. In order to
ensure the anonymity and objectivity of assessment and decrease the possibility of cheating, to simplify the
analysis of exams results Vilnius Gediminas Technical University uses the electronic testing of students. For the
creating of electronic tests IBM Authoring Tool is used. IBM Authoring Tool creates electronic tests as
SCORM-based course packages.

The process by which electronic tests are delivered to learning management system IBM Workplace
Collaborative Learning includes some stages. At the first stage the course administrator imports SCORM-based
course package to the FTP server; the Learning server imports the course package from the FTP server. At the
second stage the Learning server creates a course master and enters information about the master into the
database; course administrator registers the master and creates offerings in the course catalog. Finally, course
structure is sent to the Delivery server which registers course structure in the database, sends request to the
Learning server, and the Learning server sends the course content to the Content server [12]. The process by
which electronic tests are delivered to learning management system IBM Workplace Collaborative Learning is
showed in the Fig. 1.

![Figure 1. The process by which electronic tests are delivered to learning management system IBM Workplace Collaborative Learning](image)

The learning management system IBM Workplace Collaborative Learning uses a relational database to
store information about settings needed to run the system, courses, their structure, users, their accomplished
actions, answers to the tests, time spent for separate course activities, tests results and other interesting tracking
data. The database consists of 124 tables.

4 Research data and methods

In this research, the data of C++ Programming language exam are analyzed. For the electronic exam the
test of 25 questions generated from 100 questions set was prepared with IBM Authoring Tool. The group of
students passing the exam consisted of 100 students of full-time and evening courses from seven groups. For
every group of students the exam took place at different times, so in the learning management system, a separate
electronic exam offering was registered for every group and exam retaking. The students who did not pass the
exam could retake it two times.

All data necessary for this research were transferred to DB2 database in the local server for further
processing. IBM DWE Design Studio tool was used to perform necessary collected data pre-processing. The data
set was gathered using data selection, filtering and consolidation operations. There were created the procedures
to perform calculations. The data set gathering process is showed in the Fig. 2.

Selected data set consists of 19 fields and 3167 records. The sources of selected data are analyzable
exam records: questions given to a student, chosen answers, the times spent to answer the questions and so on.

For data analysis are chosen descriptive data mining model’s techniques: clustering and association
rules. A descriptive model which is used in this work identifies patterns or relationship in data [3]. Clustering is a
division of data into groups of similar objects. Each group, called cluster, consists of objects that are similar
between themselves and dissimilar to objects of other groups [5]. For the clustering in this work is used Kohonen algorithm. Using Kohonen algorithm the data are represented as five-dimensional vectors because the set for clustering is defined by five indicators and are divided into three clusters.

![Figure 2. The data set gathering process](image)

Another technique that describes the results is association rules. It is used to find hidden relationships. We use SIDE (Simultaneous Depth-first Expansion) algorithm that works in four steps: data examination, preparation, training and results presentation. On the initial step, the data statistics about pending data element pairs of how often it’s iterant is placed together. In the next preparation step, the most iterant data element pairs are transformed into binary format. The training step includes available rules set generation and each one rule is evaluated iteratively and chosen under suitable criteria. During next iteration excluded rules are attached to new circumstances and evaluation process is repeated again. Using this algorithm we looked for associations between e-examination data attributes. This algorithm was developed by IBM.

5 The results of data analysis

During clustering three clusters were generated. First one includes the first exam pass, i.e even 71.54% of all records; second includes first exam’s repass records – 21.95% of all; third – second exam’s retake – 6.50% of all records. Data statistics is presented in Table 1

<table>
<thead>
<tr>
<th>Cluster</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examination No.</td>
<td>Min</td>
<td>Max</td>
<td>Mean</td>
</tr>
<tr>
<td>Incorrect answers</td>
<td>2</td>
<td>21</td>
<td>12.75</td>
</tr>
<tr>
<td>Correct answers</td>
<td>4</td>
<td>23</td>
<td>12.22</td>
</tr>
<tr>
<td>Time spent</td>
<td>700</td>
<td>2556</td>
<td>1636.49</td>
</tr>
<tr>
<td>Incorrect answers time</td>
<td>198</td>
<td>1980</td>
<td>978.47</td>
</tr>
<tr>
<td>Correct answers time</td>
<td>233</td>
<td>1486</td>
<td>658.02</td>
</tr>
</tbody>
</table>

Analyzing clusters statistical data we noticed that there is deference between clustered records exam pass duration values. It is evident that students who passed the first time spent more time compared with the next passes. Although comparing with first passing time the second time was more successful, and the percent of correct answers was increased from 12.2 to 13.4, and students spent less time (about 25%).

To find out what kind of exam answers were the most difficult during first pass and first retake the association rules technique was deployed. In Table 2 the most interesting rules which were received using association are presented.

The obtained results show that students were the most serious about OOCo1q146, OOCo1q143, OOCo1q106, OOCo1q142, OOCo1q145 of questions. For example, the group of OOCo1q146 questions were incorrectly answered by even 77.80% learners. The analysis of study results highlighted the more difficult part of the course absorbed by students, refers to the course development policies. The present analysis provides
detailed feedback to the course instructor, releases should be emphasized, is difficult to understand the course is taught.

Table 2. The most difficult exam question groups

<table>
<thead>
<tr>
<th>Rules</th>
<th>Frequency</th>
<th>Confidence</th>
<th>Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>[QUESTION_ID=OOCo1q146]==&gt;[RESULT=0]</td>
<td>3.10%</td>
<td>77.80%</td>
<td>1.464</td>
</tr>
<tr>
<td>[QUESTION_ID=OOCo1q143]==&gt;[RESULT=1]</td>
<td>2.60%</td>
<td>64.80%</td>
<td>1.220</td>
</tr>
<tr>
<td>[QUESTION_ID=OOCo1q106]==&gt;[RESULT=0]</td>
<td>2.50%</td>
<td>63.00%</td>
<td>1.185</td>
</tr>
<tr>
<td>[QUESTION_ID=OOCo1q142]==&gt;[RESULT=0]</td>
<td>2.40%</td>
<td>61.10%</td>
<td>1.151</td>
</tr>
</tbody>
</table>

Continuing the investigation, we searched for the association rules of responses to difficult enough test questions between first pass and exam retake. These relationships are relevant to the analysis of student behavior during the overshoot. The selected result findings are presented in Table 3.

Table 3. The first examination and retaken results of more difficult exam question

<table>
<thead>
<tr>
<th>Rules</th>
<th>Frequency</th>
<th>Confidence</th>
<th>Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>[PASS=1]+[QUESTION_ID=OOCo1q146]==&gt;[RESULT=0]</td>
<td>1.63%</td>
<td>84.62%</td>
<td>1.5932</td>
</tr>
<tr>
<td>[PASS=2]+[QUESTION_ID=OOCo1q146]==&gt;[RESULT=0]</td>
<td>1.48%</td>
<td>71.43%</td>
<td>1.3449</td>
</tr>
<tr>
<td>[PASS=1]+[QUESTION_ID=OOCo1q143]==&gt;[RESULT=0]</td>
<td>1.41%</td>
<td>73.08%</td>
<td>1.3759</td>
</tr>
<tr>
<td>[PASS=2]+[QUESTION_ID=OOCo1q143]==&gt;[RESULT=0]</td>
<td>1.19%</td>
<td>57.14%</td>
<td>1.0759</td>
</tr>
</tbody>
</table>

The obtained results show that students during the exam retaking answered tougher test questions more accurately. Interpreting the findings suggest that the students examined were better prepared to retake or knew the right answers to the questions provided in the retake.

Continuing the analysis we searched for association between time spent and the right answer. First of all, we analysed the first test data. To find associations exam time spent is broken down into 12 intervals. The obtained results are presented in Table 4.

Table 4. The relationship between time spent and correct answer during first pass

<table>
<thead>
<tr>
<th>Rules</th>
<th>Frequency</th>
<th>Confidence</th>
<th>Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>[DURATION&gt;=5 AND &lt;22.5]==&gt;[RESULT=0]</td>
<td>8.15%</td>
<td>68.62%</td>
<td>0.8167</td>
</tr>
<tr>
<td>[DURATION&gt;=5 AND &lt;22.5]==&gt;[RESULT=1]</td>
<td>8.62%</td>
<td>31.38%</td>
<td>1.2698</td>
</tr>
<tr>
<td>[DURATION&gt;=22.5 AND &lt;40]==&gt;[RESULT=0]</td>
<td>16.15%</td>
<td>58.99%</td>
<td>0.9908</td>
</tr>
<tr>
<td>[DURATION&gt;=22.5 AND &lt;40]==&gt;[RESULT=1]</td>
<td>11.23%</td>
<td>41.01%</td>
<td>1.0136</td>
</tr>
<tr>
<td>[DURATION&gt;=40 AND &lt;57.5]==&gt;[RESULT=0]</td>
<td>12.31%</td>
<td>61.54%</td>
<td>1.0136</td>
</tr>
<tr>
<td>[DURATION&gt;=40 AND &lt;57.5]==&gt;[RESULT=1]</td>
<td>7.69%</td>
<td>38.46%</td>
<td>0.9506</td>
</tr>
<tr>
<td>[DURATION&gt;=92.5 AND &lt;110]==&gt;[RESULT=0]</td>
<td>4.92%</td>
<td>59.26%</td>
<td>0.9953</td>
</tr>
<tr>
<td>[DURATION&gt;=92.5 AND &lt;110]==&gt;[RESULT=1]</td>
<td>3.38%</td>
<td>40.74%</td>
<td>1.0069</td>
</tr>
</tbody>
</table>

The table shows that the correction of a given answer depends on time. The students who were spending on the response from 5 to 22.5 seconds answered test questions correctly most often. If the candidates answering the question spent more than 180 seconds, the answers are often not correct.

In finding associations between the time spent to answer a question and correct answering during first-time retaking very similar results were obtained. Students, who respond within 40 seconds, usually answer the question correctly while those who take more than 180 seconds often were wrong.

Analyzing the reasons influencing these results, we can accurately identify and formulate the incidence of ambiguous exam questions. Therefore, the exam questions answering by students with difficulty and taking the most time should be further reviewed and revised.

6 Conclusions and further work

For analysis of electronic exam data were chosen descriptive data mining model’s techniques: clustering and association rules. They help to identify patterns and relationship in electronic exam data.

After clustering technique we obtained statistical data which showed that the behavior of each student is different - each test retaking decreases the mean of exam time duration and increases the mean of the number of correct answers.
Using the association technique we examined questions that were given during the exam and student responses to them, received a set of rules, representing the most complex sets of questions. Due to the association technique the rules which showed students knowledge have been found.

The obtained results show the behavior of students in each test, as well as the exams, their reliability and the benefits of the study process.

In this work we suggest to apply data mining techniques to find the appropriate e-learning patterns, related to student evaluations and time spent for answers. This work will be used to improve VGTU electronic examination system.

The small-scale research of e-examination results was carried out, using data extraction techniques, and it presented an interesting and useful knowledge, according to which it is possible to establish a higher quality e-tests and to improve the examination process. However, this study used relatively limited data, the number of questions to which answers we had hoped for has not been great.

Further we plan to analyze the distance learning information and e-examination results of Master students. Currently we have the data for e-learning and e-examination data of two subjects (Virtual learning environment and Multimedia) for three years. Each year, about 20-30 graduate students study the courses. In e-environment they are not only examined, but are also supplied with all the course materials; they can get consultations, practical exercises, and self tests are provided. During this study, we will seek to identify more factors influencing the successful learning practice.

References


DISTRIBUTED ARCHITECTURE FOR CONTEXT MODELING BASED E-LEARNING SYSTEM

Kazys Baniulis, Giedrius Paulikas, Jurate Pauliute, Ruslanas Sobolevas, Gytis Vilutis

Kaunas University of Technology, Department of Computer Networks, Studentų 50, Kaunas, Lithuania, kazys.baniulis@ktu.lt, giedrius.paulikas@ktu.lt, jurate.pauliute@ktu.lt, ruslanass@gmail.com, gytis.vilutis@ktu.lt

Abstract. Simulation-based e-learning allows students to use models imitating the real processes and thus accumulate knowledge and experience acquired in the process of active learning. This paper presents the advanced architecture of virtual interactive e-learning system TestTool that is under development in Kaunas University of Technology. TestTool system is implemented as an assessment and graphic modeling engine and is based on architecture of distributed services that is expected to transform to service oriented architecture in the future. The system consists of operation center and resource repository services that are accessible through admin, author and student client programs. In particular, context models are used for construction and assessment of graphic models that are available during learning process. Context models are created using additional design module that can be used as TestTool subsystem. These models describe the subject domain by using contextual graphs and feature diagrams. The employment of context model helps experts to simplify and speed-up the creation of effective graphic models for any given subject and to ensure proper and just-in-time up-/reskilling that can be used in continuous process of education and training.

Keywords: e-learning technology, active learning object, assessment, context modeling, contextual graph, distributed architecture.

1 Introduction

The quality of e-learning technologies of higher and secondary education, corporate learning and vocational training must be the main factor that governs their expansion [10]. E-learning should be developed for encouragement of widespread learning, cheaper learning and acquisition of the latest and most innovative material. More flexible coherence of the study curricula with student needs and individuality, creation of opportunities to avoid emotional tension and ensuring more objective assessment of all learners is also very important [10]. Thus, good results of studies depend on many factors of learning motivation, where the first premise is the change and dynamism of life. Learning technologies should adequately reflect the content of changes that come from the experience of dynamic life.

In advanced learning technologies the instructional material (content) is composed of Learning Objects (LO). Learning Object can to be seen as a logical container that represents an atomic web-deliverable resource such as Lesson (HTML page), Simulation (Java applet) or Test (HTML page with evaluation form). A high quality instructional material must contain the expository materials, active learning and (self-) assessment parts that constitute the full cycle of experiential learning [13]. The implementation of active learning is the most sophisticated part of e-learning technology design.

This paper discusses how simulation, experimentation and practice as well as assessment and self-control features can be implemented in active learning objects (ALOs). Simulation, practice and experimentation tools are usually created with Java, Flash or similar technologies that have high costs of development. We will show that very effective and fast applications for (self-) assessment in simulation-based active learning can be created for any particular domain by making the information model of the domain that later is used to design and implement the situations of interest.

2 Using context modeling for ALO design

This chapter discusses ALO peculiarities and its implementation that uses context modeling [7][9][17]. The important notions of context and contextual element will be presented. The new ALO information model will also be discussed as it helps to design better ALOs and forms the means for generation of explanations.

2.1 Active learning object

Learning and education applications that rely on modern information and communication technologies are oriented to active studies that are implemented by virtual learning environments through various methods of active learning: tests of knowledge assessment, modeling practice, practical tasks, experimental tasks, interactive presentations, etc. The applications of active e-learning are based on experiential learning model of Kolb, where
it is stated that “Learning is the process whereby knowledge is created through the transformation of experience” [13]. This model emphasizes the appreciation and usage of already accumulated experience to solve new problems and lends this ability to active learning which allows its user to master new and increasingly complex situations through references to already familiar ones [13]. The educational content is decomposed to separate problems that are solved by learners relying on their experience as well as new learning material. Active e-learning isn't linear or consecutive, i.e. the students are not instructed to absorb the content or solve problems step by step. They rather make decisions that are based on content, because, as noted by Baniulis and Tamulynas expository content presentation and assessment tests are not sufficient for comprehensive evaluation of acquired skills [4][19]. Many researchers (Bauer [2], Baniulis and others [4][19]) found that learning through activity is far more efficient, as modeling practice stimulates the active behavior and constructive learning process. E-learning environments tend to use more modeling practice that focuses students on problem at hand instead of separate fragments of domain knowledge. But lack of feedback in this learning method inhibits motivation and active behavior [15]. To solve this shortcoming Bauer suggests using the feedback link that consists of knowledge evaluation which is included in the modeling itself [2].

Among the most important features for effective learning materials is the level and type of interactivity, understood as the dialogue between students and presented materials or systems. Various levels of interactivity and difficulty are needed in active LOs, that offers users the opportunity to adapt the presentation of the content to their individual skills and needs, thus securing fast, situated, just-in-time up-/reskilling that is used in continuous education and training processes. It's important to ensure the flexible design of ALOs and comprehensive generation of user explanations. In order to achieve these goals the paper proposes a new information model of ALO that is based on context modeling. It relies on Brezillon research of Contextual Graphs (CG) and ALO information model research of Slotkienè [17]. The following is the discussion of conception of context that lies at the base of creation of context model.

2.2 Context

Traditionally according to LOM specification, context is the environment in which the learning model (or learning object) is functioning. It is described by several parameters of metadata that characterize the environment. But in reality context is very complex. Context modeling in particular [9][17] has the notion of context that is much wider, it contains both the model (ALO) context and its environment context. Context has data and information [9] that describes elements of the domain, their relations and regularities of their changes. This way context can be viewed as a systematic description of the domain. In case of this conception of context we can implement the systematic approach that is opposed to the partial and one spot approach to the phenomena, e.g. common in traditional tests. In traditional tests, questions are often designed to contain one particular piece of knowledge; the answer to each question is seen as an independent data point [2]. In contrast, the individual actions within sequence of interactions of a simulation are often highly dependent on one another. Creation of graphic construct model by means of context modeling implements the methodical approach to phenomena that takes into account interaction among the elements of the whole system which constitutes an integral entity [3].

Context, as description of the part of the domain, is complex and dynamic, it covers various situations. The foundation of context information model (IM) is made of units that represent solutions of elementary tasks (solution steps), they that are called contextual elements.

2.3 Contextual element

Contextual element (CE) can be broken to focus, contextual data, rule and information [7][9]. Let's examine a simple example.

Let’s say the student already knows the structure of the process for cyclic calculations and he/she has to understand the algorithm for cyclic sum of array elements and find out the differences from product algorithm. Let’s take a simple problem of initial value of the accumulating variable. When the task means that algorithm = sum, the student must assign the accumulation variable initial value of 0, but when algorithm = product, the assigned value should be 1. Moreover, correct/incorrect are not sufficient as feedback of learning process, exhaustive explanations should be provided why the chosen solution was right or wrong and where the mistake was made.

Figure 1 shows a model of one CE. The sample ALO for algorithm of cyclic calculations is on the left, it contains the task and the solution with two empty spaces. The student has to fill in two sentences of this algorithm. The information model of filling the first sentence is presented on the right of the block diagram: i.e. focus, the possible data (Data 1) and result (Data 2), information and rule. A sequence for applying the rule to the particular task and getting the result is also given. The sequence of actions for filling in the first sentence of the algorithm is shown in Table 1.
Table 1. The sequence of CE actions

<table>
<thead>
<tr>
<th>Action</th>
<th>Name</th>
<th>Content</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Focus</td>
<td>Context analysis (Task 1 and graphic model)</td>
<td>Focus F1</td>
</tr>
<tr>
<td>2</td>
<td>Information</td>
<td>Analysis of Task 1 and Data 1</td>
<td>Information  (\text{algorithm} = \text{sum})</td>
</tr>
<tr>
<td>3</td>
<td>Rule</td>
<td>Rule input (\text{algorithm} = \text{sum})</td>
<td>Initial value (S = 0)</td>
</tr>
<tr>
<td>4</td>
<td>Action</td>
<td>Deduction: (S = 0) must be placed to focus F1; Activity: place sentence (S = 0) to focus F1</td>
<td>Sentence (S = 0) in focus F1</td>
</tr>
</tbody>
</table>

**Alternative:** The same CE model can have different task (Task 2) where focus F1 in the algorithm of array element product must be filled with sentence \(S = 1\).

Information model that is shown in Figure 2 has two parts: the upper FD and lower CG. CG enumerates the tasks and their solutions that are available in the chosen topic. CG has two types of nodes: context elements and actions. Context element describes a step in the solution of the task, while the whole process of solution is presented by the graph path. The example of information model for cyclic calculations is given. The variability of topic data is reflected in FD: cyclic calculations can have three types of algorithms for sum, product or count. The initial value of accumulation variable can be 0 or 1 and it can be increased by the supplied value or 1. The data is related, e.g. accumulation variable is increased by 1 in the algorithm of count calculation.

In Figure 2 CG illustrates the example given in Figure 1, therefore it describes only a part of FD. CG describes the three possible tasks of the graphic model: the given block diagram can be filled in to create algorithm of sum, product or count. The first step of solution, filling in the initial value of accumulation variable, is described by CE C1. If task is to form the sum algorithm, then after checking the condition of context element C1, \(\text{Is algorithm} = \text{product}\)?, we get a negative answer, and, after choosing side branch C1_N, perform action A1 (\(S = 0\)). The next branches are chosen by answering other questions of the CG path and lead to the actions of the problem solution. Construction of graphic model from contextual graph allows to verify the correctness of solution actions, specify errors and form explanations for the entire solution, where explanations gives details about the correctness of the solution and reveal the places of errors (if any). IM of CG can be used to form different ALOs that encompass one or more CEs, but ALO (graphic model) must include all paths that start at its CE. If ALO is made of a small number of CEs, there will be little paths that lead from them and, consequently, little variants (ALO instances). Such ALO will be easy (low level of difficulty). In order the number of CEs increases, the ALOs will vary in difficulty.

### 2.4 Context model based implementation of ALO

Implementations of modeling practice employ many means ranging from specialized software to explicitly authored scenarios. The graphic system of distance assessment TestTool is under development in Kaunas University of Technology for almost a decade and is successfully used for modeling practice. The main feature that separates TestTool system from other similar tools is the ability for learners to perform practical actions that are required to answer graphic questions. This assessment system allows constructing various problematic situations that are later presented for solution. This way student constructs the solution rather than selecting it.
Let's discuss the learning environment of some small or medium enterprise. Enterprise N collects from clients their faulty computers, an expert inspects them by running various tests, observes and analyzes test results, changes faulty components and returns computer to their owners. Construction of learning environment from the processes of computer repair that are used in enterprise N starts with creation of IM (made of FD and CG). TestTool uses IM to form the graphic model of computer repair (ALO). The base of this graphic model is the graphic pattern of testing scenarios that is prepared in accordance to CG. The creation of graphic model was supported by clips of real documents, screenshots of testing programs and other practical material. Four different versions of graphic model were created and each version was used to construct easy, moderate and difficult ALOs accordingly to the requirements of CG and FD.

Figure 3 shows a version of graphic model, where the presented context is made of the data of computer inspection, the graphic pattern of testing scenarios and the lists of possible conditions (C) and actions (A). For the given task student can use fault description *No power (no signs of life)* to construct the computer testing scenario, while modeling system will evaluate correctness of the solution and describe the errors. In the graphic pattern of testing scenarios student has to select from the list only those Ci and Ai that are required to repair the given computer. Computer faults in other ALO tasks will have different fault descriptions or testing screenshots, they will require different actions for inspection and component testing/replacement, i.e. we will have various scenarios of computer testing and repair.

The earlier described graphic model is created by means of graphic testing. In order to formalize the structure of design process, fully implement features of modeling-simulation and achieve higher degree of
system flexibility, a new version of the system is developed. The new version of TestTool 5.3 implements modeling-simulation features that use the new context modeling based information model of ALO (see chapter 2.3). The system was also redesigned to support improved flexibility and upcoming transition to service-oriented architecture. The next chapter discusses the details of the initial phase of this implementation – the architecture of TestTool 5.3.

3 Architecture of TestTool 5.3.

The TestTool 5.3 system is implemented as an assessment and graphic modeling engine. It’s a modular client-server system with several classes of users. The current system works as a network of distributed services and is planned to refactor to service-oriented architecture. Version 5.3 already supports the flexible incorporation of new scenarios and overall system changes.

3.1 Structure and functions of the system

The system consists of resource repository (RR) and operation center (OC) services that are consumed by administration, student and authoring client software (Figure 4). TestTool system can be used by users of three different levels – super-administrators, administrators and learners. Learning resources are made of courses, exams, tests, questions, variants and question answers. The educational process consists of the preparation of learning material (initial preparation stage) and its application for training of student groups (learning stage). During the preparation stage question variants are created (or reused) and combined to questions, tests and exams (administrator/author responsibility). Learning stage involves students that prepare for the assessment (exams are in practice mode) and later take tests (exams are in assessment/testing mode).

As already mentioned, there are several different roles of TestTool 5.3 actors/users. Super-administrators use OC GUI to manage the descriptions of administrators. Administrators employ the same GUI for registration of courses, users, user groups, tests and exams; they also select the evaluation scenarios. When acting as a domain expert, administrator uses authoring tool GUI to create questions and implement knowledge structure of graphic models. Students have the student GUI that is used to connect to the system, select working mode and take various tasks. In practice mode learner can freely choose tasks and take them multiple times, but results of these actions are not saved. In testing mode each task of the exam can only be taken once and all results are saved in RR.

Operation centre has access to resources in RR and carries out administration of users and tests, it also performs user authentication and checks execution rights for all operations. Access rights are also transferred from OC to resource repository that has no other means of data protection (see figure 4). A single operation centre can use several resource repositories that are accessed in server-client mode. RR service stores questions, student answers to those questions and tools for general resource management. RR uses special scenarios to evaluate the answers. Resource repositories are managed by administrators through GUI of operation centre. Each administrator can be assigned more than one resource repository and RR can belong to several administrators.

The authoring tool is used by administrator, or domain expert, to implement the knowledge structure of graphic model. This authoring tool itself contains many technical features that can be used for that purpose: drawing primitives, elements for simulation control, textual and graphic explanations of errors, metadata and various multimedia effects. It facilitates the creation of active learning objects with different levels of interactivity [4]. Author tool is used for creation of both traditional questions and models of graphic construction (they are coordinated with the contextual graph). These question types also need different scenarios of answer evaluation. The authored questions can be directly uploaded to the selected resource repository or they can be
saved to hard disk and later uploaded through OC. The RR stored learning material can be private for the administrator that uploaded it but also can be made public to all administrators that use the services of that RR.

3.2 Scenarios and their application

The system employs different scenarios for preparation and execution of studies. They facilitate the implementation of flexible structures for assessment and graphic modeling. Arbitrary number of scenarios can be added as system extensions in Java Archive files (*.jar). Figure 5 shows a sample course with 2 exams. Exam2 with traditional questions and graphic model simulation tests in practice mode is created for preparation, Exam1 is for assessment. Each exam is composed of one or more (two in this case) tests that comprise a set of questions. Exams and tests are assigned evaluation scenarios that govern how student answers are treated. Tests have two scenarios: one for questions (different scenarios are required for traditional and simulation questions) and one for the whole test. Exams need just one scenario that sums test results.

The preparation phase covers resource editing and making learning material (exams) ready for studies and assessment. To construct a new exam administrator has to carry out these actions: 1) create exam name, 2) select private (personal) or public question from the storage, 3) create test and assign questions to it, 4) assign test to the exam, 5) choose the scenario for presentation of question variants, 6) choose the scenario for evaluation of questions or model, 7) choose the scenarios for test and exam evaluation, 8) assign student group to the exam. During preparation phase resources can be edited according to the patterns of preparation scenarios.

When learning phase is active, student program uses OC to get exam questions from RR, put results of question answers to RR and receive evaluation from RR. Scenarios for learning phase support:

- Question evaluation – evaluates answers to traditional questions;
- Model evaluation – evaluates answers to questions with graphic models according to information model CG;
- Test or exam evaluation – concludes the evaluation of test or exam, e.g. test questions can be evaluated as weighted or not, "exam evaluation" scenario can use some expression to calculate the final evaluation of several exams, etc.;
- Presentation of question variants – designates the number of question variants that are presented to the student, the way of variants selection (e.g. random or metadata driven user choice), etc.

Learning material of traditional tests is meant for preparation and self-studies, usually it can't be applied to assessment, in which case extra protection is required. Thus the ability to use different modes of same questions (Test5 in Figure 5) both for preparation (Exam2, simulation mode, in Figure 5) and assessment (Exam1 in Figure 5) is very important.

3.3 System implementation

Resource repository and operation centre are implemented using the same programming technologies: Java objects, relational database and design patterns that have become popular in the domains of architecture, software design, human computer interaction, Web 2.0, organizational structures and pedagogy as a way to
communicate successful practical knowledge. Patterns capture proven solutions for recurrent problems with respect to fitting contexts [12]. The standard enterprise Java pattern of Data Access Object (DAO) [1] was used to clearly separate the business logic and database layers that improved the maintainability and extensibility. DAO design pattern supports the object persistence and data access logic independence from any particular persistence mechanism or API. This approach provides the flexibility to change application's persistence mechanism over time without the need to re-engineer application logic that interacts with the DAO tier [19].

TestTool system provides a simple, consistent API for data access that does not require knowledge of specific technologies, thus widening the base of extension and plugin developers. All database access in TestTool 5.3 system is made through DAO that supports data encapsulation. Each DAO instance is responsible for the single primary domain object or entity, but only if a domain object has an independent lifecycle. At the lowest layer Hibernate is used to connect application and database, it's an excellent ORM framework which allows developers to access database by using familiar object-oriented approach [19]. Hibernate has the following characteristics: transparency in providing mapping between object and relational database, support for multiple databases by using a unified interface, advanced caching mechanism and locking strategy that significantly reduce operations against the database. It's provided under free open source license; source code can be consulted and modified for required functional customization. Hibernate is a lightweight package, it doesn't introduce many complex constructs, so it's easy to debug, that reduces the burden of TestTool maintenance and extension programmers [20].

The whole TestTool system is implemented as a set of program layers; each of them has clear boundaries and interconnections to adjacent layers. Layered structure is essential for evolving system that is constantly extended for research of new e-learning features and adapted to various middleware platforms.

4 Conclusions and future works

1. Creation of high quality e-learning technologies that simulate sophisticated and dynamic systems requires the simplification of learning models by decomposing them to smaller problems that can address individual skills and needs of the learners. The paper presented a new information model that comprises contextual graph and a feature diagram, this model is taken as a base for implementation of experimental tool for ALO construction TestTool 5.3. Simulation, experimentation and practice as well as explanation, assessment and self-control functionality can be implemented using active learning objects (ALOs).

2. ALOs are formed from CG sub-graphs with different number of CE nodes and all paths that lead from these nodes. This creates the ability to consistently change ALO difficulty while ensuring completeness and adequacy of the model. Course with such ALOs (i.e. LO of higher aggregation level) fulfills the completeness requirement by realizing all phases experiential learning cycle.

3. Very effective and fast applications for (self-) assessment and simulation-based active learning can be created for any particular domain. Real world usage of ALO is reinforced by utilization of authentic materials. Noteworthy is the possibility to use tasks with ALO graphic models both in practice and testing modes, it clearly distinguishes such tasks from the traditional tests.

4. The TestTool 5.3 system is implemented as an assessment and graphic modeling engine. Future works include further development of TestTool subsystem for context model design, implementation of CG XML parsing in author and student programs, improving generation of explanations, implementation of standard compliant learning services and transition to service oriented architecture (SOA).

5. The second stage of development includes plans for transition to SOA with reusable services. Grid middleware based distributed service architecture was already successfully used in earlier versions of TestTool [19]. SOA implementation is expected to support the following features:

• High level of abstraction where business logic is completely hidden from service consumers. Implementation of this feature is facilitated by Date Access Object and logic tiers that hide main business logic and expose only RMI interfaces.

• Loose coupling: TestTool services have minimal interdependencies and are ready for refactoring to web services.

• Service statelessness: Resource Repository stores minimal amount of data context information so any invocation of this service is not bound to earlier or later service invocations.

Resource repository and operation centre can be easily implemented as two individual services that satisfy main requirements of SOA. The upcoming version of TestTool system can be created as EJB container that is accessible through RMI and web services (only RMI interface is available now). Having two separate interfaces are useful for situations where better performance of RMI or standard compliance of web services are required.
References


