USING STORYBOARDING AND DATA MINING TO ESTIMATE SUCCESS CHANCES OF CURRICULA

Rainer Knauf, Ronald Böck
University of Ilmenau, Faculty of Computer Science and Automation
PO Box 100565, 98684 Ilmenau, Germany
Rainer.knauf@tu-ilmenau.de

Yoshitaka Sakurai, Shinishi Dohi, Setsuo Tsuruta
Tokyo Denki University, School of Information Environment
2-1200 Musai-gakuendai, Inzai, Chiba, 270-1382, Japan
{ysakurai,dohi,tsuruta}@sie.dendai.ac.jp

ABSTRACT

In university studies, there is a flexible but complicated learning system of subject offers, enrollment rules for particular subject combinations, and prerequisites to meet for taking particular subjects, which need to be matched with students’ needs and desires. Students need assistance in the jungle of such learning opportunities and limitations at today’s universities. To face this problem, we employed our formerly developed storyboard concept and used it to develop, maintain, and evaluate curricula. Storyboarding is based on the idea of formally representing, processing, evaluating and refining didactic knowledge. This concept is utilized to supplement an educational system called Dynamic Learning Needs Reflection System (DLNRS) of the School of Information Environment of Tokyo Denki University, Japan. Concretely speaking, didactic knowledge of DLNRS can be represented by storyboarding and used for supporting dynamic learning activities of students. Here, we introduce an additional benefit of the storyboard concept. By using data mining-like methods to evaluate storyboard paths, we are able to estimate success chances of storyboard paths. Based on such an evaluation we will be able to rate planned (future) paths and thus, to prevent students from failing by non-appropriate curricula. Moreover, besides the evaluation, the estimation can be used for computer enforced suggestions to complete a path towards optimal success chances.

KEYWORDS

Process modeling, Learning processes, Storyboarding, Dynamic learning need reflection system, Success estimation

1. INTRODUCTION

University studies are characterized by a high degree of flexibility with respect to the subjects to be included. On the other hand, there is a complex regime of subject enrollment rules, requests on the methods and time of subject enrollment, and prerequisites that needs to be met for entering particular subjects. Under these circumstances it is not easy to finish a study in time and with success. Both the flexibility and the complexity of regulations increase the study period enormously. The authors’ experiences with Japanese, German, and US universities indicate that this is a world wide general problem.

Generally, university studies suffer from a deficiency of clarity due to the above mentioned regime. An unacceptably high number of students fail because they couldn’t comply with some of these regulations or didn’t even know them. Also, many students can’t finish their study in the designated time for the same reason. Avoiding the resulting frustration is one objective of introducing storyboarding as a means to keep the overview in the jungle of opportunities and restrictions. Typically, academic education at universities is characterized by (1) a large variety of opportunities to compose academic time lines or curricula and (2) teachers (professors and tutors), which are usually excellent experts in their subject, but do not necessarily have the didactic skills to teach their subject.

In particular, at the School of Information Environment (SIE) of Tokyo Denki University (TDU), today’s students are required to be more flexible in designing their study according to their needs, wishes, interests,
and talents. To meet this request, an education system, which we currently call Dynamic Learning Needs Reflection System (abbreviated as DLNRS), has been developed and introduced at SIE of TDU (Dohi & Nakamura, 2003). Its objective is to keep and increase the students’ motivation through clarifying and dynamically reflecting students’ learning needs by themselves. The system is characterized by (1) the abolition of the traditional rigid academic year, (2) the introduction of prerequisite conditions instead of a fixed pre-determined subject sequence, (3) the displacement of a fixed charge per year by a subject-oriented paying system, and (4) a Grade Point Average (GPA) system to rate the learning results and to derive appropriate consequences for the upcoming educational process schedule at TDU.

Qualified guidance needs adaptation, i.e. dynamics with respect to varying needs, context conditions, the student’s performance, and the students’ educational history. Adaptation, however, presumes an anticipation of different alternatives and their explicit representation. Didactic variants have to be subject to discussion and quality assurance. For this purpose, an appropriate didactic design practice needs to be established. Didactic design means the anticipation of those communication processes (Flechsig, 1996), and storyboards may provide the expressive power suitable to the design and implementation of learning processes.

Storyboarding as introduced in (Jantke & Knauf, 2005) and (Knauf et al. 2007) is a very general concept. In the context of DLNRS, storyboarding complements the system so far. Through adopting the storyboard concept for a complete university study, also the management of the study becomes accessible for evaluation and refinement, i.e. quality assurance. As a deeper benefit of this work some data mining can be performed over the paths of particular students after they completed their study at TDU. This will answer questions like “What do the successful students’ paths have in common?” or “What do their paths distinguish from the ones of less successful students?” and finally make the students able to create optimal curricula.

Based on such technologies, we recently developed an approach to evaluate curricula created or modified by the students in advance of their study (Boeck, 2007). The basic idea is twofold. It consists of (1) the construction (and successive refinement) of a decision tree based on paths that have been followed by (former) students, i.e. path with a known level of success and (2) its application to estimate the possibility of success (success chance) of a planned path, where (future and current) students want to go.

The tree construction is based on a “flatten” storyboard, i.e. a huge storyboard that is constructed from the top level storyboard by replacing each episode by its related sub-graph at all hierarchy levels. The tree forks at nodes with different successor node. The tree’s leaves are the (known) success level (examination results).

The estimation of success chances for a newly built schedule is performed though traversing the tree until the path contains a “next node” different from all successors of the related tree node. The estimated success level is computed as the weighted average mark of the sub-tree beginning at this node. Moreover, (Boeck 2007) introduces a technology to suggest a modification to a given curriculum that leads to an optimum with respect to the success chances.

Before we show the way to adapt the storyboarding concept for the intended application, a short introduction to the DLNR system needs to be briefly introduced in the next section for better understanding.

The paper is organized as follows. Section 2 describes the DLNRS as successfully introduced at the SIE of TDU. Section 3 is an introduction to the storyboard concept as developed so far. Section 4 introduces the utilization of storyboards for the estimation of success chances for curriculum plans.

2. DYNAMIC LEARNING NEED REFLECTION CONCEPT

In its concept, DLNRS (Dohi & Nakamura, 2003) primarily aims at promoting the students’ motivation by creating or modifying their own class schedule per semester or graduation time lines by themselves. This is a way to develop a spirit of independence and to keep up with globalization. Key features of DLNRS are:

1. Abolition of the traditional rigid academic year
   There is no academic year with fixed courses and a fixed fee. Instead, there is a semester-based course system with a tuition fee for each particular subject. There is no restriction for attending a particular subject in a particular semester besides the prerequisites specified for this subject. Thus, the students are able to study at their own adaptive paces.

2. Abolition of compulsory subjects
   Specific compulsory subjects have been replaced by the concept of prerequisite conditions. These conditions are expressed in two levels of recommendation (1) subjects that have to be learnt before
and (2) subjects that are recommended to be learnt before. The prerequisites are formally checkable by considering the Grade Points received in the subjects that are prerequisites.

3. Replacement of a fixed charge per year by a subject-oriented paying system

Students pay a subject-oriented fee according to the number of units of the subject. Therefore, they carefully check their learning needs to pick out the right subjects to achieve their academic goal. This motivates to make a maximum effort to pass the subject with respect to the money invested.

4. Class period length

The usual length of a class is cut down from 90 min to 50 min and 75 min. Typically, a subject is taught in 3 units either as $3 \times 50$ min or as $2 \times 75$ min a week. The intended effect is that students will be able to concentrate the entire length of a class. Therefore, it is a contribution towards more learning benefit from the subjects and thus, from the money spent for it.

5. Grade Point Average (GPA)

This is a system to rate the learning results and to derive appropriate consequences for the upcoming educational process schedule. The GPA of subjects is calculated by equation (1) with $g_i$ being the points earned for a particular subject, $u$ being the number of units of this subject, and $n$ being the number of subjects in the semester. The number of Grade Points ($GP$) per subjects ranges from 4 (> 80%) down to 0 (< 40%).

$$\text{GPA} = \frac{\sum g_i}{\sum u_i}$$

The intention of this measure is that the maximum number $n_{max}$ of units a student can register is controlled by the GPA of the previous semester as shown in equation (2). The latter regulation is a consequence from the experience with students, who are obviously not able to self-estimate their capacity. In the trade-off between (a) a high learning quality, which is indicated by a high GPA, and (b) a high learning quantity, which is indicated by a high number of units, some students tend to promote the latter at the cost of the first.

The introduction of the DLNRS at the SIE is supported by (a) a Curriculum Planning Class, which aims at developing an individual curriculum for each student by himself that meets his needs and desires and (b) a Workshop, which aims at developing an ambience of mutual trust between a professors and his students.

Since the relationships between (1) the prerequisite conditions, (2) the GPA, (3) the quantitative unit composition regulation for graduation and (4) other aspects are difficult to overview, the development of class schedules and long term graduation timetables is a quite challenging task. Therefore, we supplemented the DLNRS concept through the concept of storyboarding.

In the Curriculum Planning Class, students who have enrolled with SIE create their graduation timeline. The completed graduation timeline is submitted to the school by e-mail, including post-graduation goals. While this may seem to be a daring plan, it is natural to make clear each student’s individual target upon entering the university. Here, the suggested data mining method can be used to estimate the success chance of a submitted plan. Moreover, supplements can be proposed, which improve the plan towards an optimal success chance.

For better understanding, we continue with a section on the storyboarding concept in general. After that, a separate section is dedicated a way to utilize the synergy of both concepts for the intended purpose.

3. STORYBOARDING

The storyboard approach adopted here (Jantke & Knauf, 2005; Knauf & Jantke, 2006; Knauf et al., 2007) is built upon standard concepts which enjoy (1) clarity by providing a high-level modeling approach, (2) simplicity, which enables everybody to become a storyboard author, and (3) visual appearance as graphs.

A storyboard is defined as follows:

- **A storyboard** is a nested hierarchy of directed graphs with annotated nodes and annotated edges.

- **Nodes** are scenes or episodes. **Scenes** denote leaves of the nesting hierarchy and represent a non-decomposable learning activity, which can be implemented in any way. It can be (1) the presentation of a (media) document, (2) the opening of any other software tool that supports learning (e.g., an URL and/or an e-learning system) or (3) an informal description of the activity. There is no formalism at and below the scene level. **Episodes** denote a sub-graph.
Graphs are interpreted by the paths, on which they can be traversed. There is a Start- and End-node for each graph. The Start Node of a (sub-) graph defines the starting point of a legal graph traversing. The End Node of a (sub-) graph defines the final target point of a legal graph traversing.

- Edges denote transitions between nodes. The rules to follow an edge are (1) the outgoing edge must have the same color as the incoming edge by which the node was reached and (2) if there is a condition specified as the edge’s key attribute, this condition has to be met for leaving the node by this edge.

- Nodes and edges can carry key- and/or free attributes. Key attributes of nodes specify application driven information for all nodes of the same type, e.g. actors and locations. Key attributes of edges specify conditions, which have to be true for traversing by this edge. Free attributes may specify whatever the storyboard author wants the user to know: didactic intentions, useful methods, necessary equipment, e.g.

The interpretations of these terms are described after presenting a small example.

Figure 1 shows a top level storyboard that models the anticipation of the diverse ways to study a recently submitted paper according to the readers’ individual purposes. The sections of the paper that is currently under a reader’s (your) consideration appear as the storyboard’s episodes, if they have a substructure and as its scenes, if they don’t.

Further structured sections are Episodes (with subsections). They need to be implemented by constructing a related sub-graph. Episodes are represented by a rectangular with double vertical lines.

Each Episode is followed by a (pentagonal) reference node, which is the re-entry point into the graph after reaching the End-node of the sub-graph for the Episode. Sections with no further structure are scenes without subsections. They are represented as rectangles. If a Scene does not really introduce new topical content (like the reference list, for example), it is represented by an ellipse.

The representation as a graph (instead of a linear sequence of sections) reflects the fact that different readers trace the paper in different manners. According to their particular interests, prerequisites, current situation (like time pressure, e.g.), and other circumstances. The various alternative paths to study this paper, for example, may be driven by the reader’s role as follows:

- Members of Ilmenau research group may skip the Introduction and Summary and Outlook sections as well as the section on the Storyboarding concept, because they are familiar with it.
- Members of the Tokyo research group may skip the Introduction and Summary and Outlook sections as well as the section on the Dynamic Learning Need Reflection System, because they are familiar with it.
- Referees may (hopefully) want to read all sections. After reading the Summary and Outlook section, they can read the Acknowledgements and References independently from each other (in any sequence). They don’t have to read the Acknowledgements, but for their duty they have to read the References at least.

A storyboard can be traversed in different manners according to (1) users’ interests, objectives, and desires, (2) didactic preferences (e.g. the need of examples or illustrations to better understand), (3) the sequence of nodes (and other storyboards) visited before (i.e. according to the educational history), (4) available resources (like time, money, equipment to present material, and so on) and (5) other application driven circumstances.

In fact, the storyboard is a semi-formal knowledge representation for the didactics of a teaching subject. Thus, it is effective as a firm base for processing, evaluating and refining this knowledge. The vision of this idea’s further effect is to gain didactic knowledge by analyzing storyboard paths by means of Data Mining methods. The node types, their visual appearance, their behavior on double click, and their behavior when following a hyperlink are as specified in Tables 1-5.
### Table 1. Scene

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Behavior when double clicked</th>
<th>Behavior on following a hyperlink</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• nothing, if just verbally described scene</td>
<td>• visiting a website with the standard browser, if it is an URL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• opening the standard mail tool, if it is an e-mail address</td>
</tr>
</tbody>
</table>

### Table 2. Episode

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Behavior when double clicked</th>
<th>Behavior on following a hyperlink</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>• visiting a website with the standard browser, if it is an URL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• opening the standard mail tool, if it is an e-mail address</td>
</tr>
</tbody>
</table>

### Table 3. Start Node

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Behavior when double clicked</th>
<th>Behavior on following a hyperlink</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>jumping to the Start Node of the related super-graph</td>
<td>not meaningful</td>
</tr>
</tbody>
</table>

### Table 4. End Node

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Behavior when double clicked</th>
<th>Behavior on following a hyperlink</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>jumping to the Reference Node that successes it’s associated Episode Node in the related super-graph</td>
<td>not meaningful</td>
</tr>
</tbody>
</table>

### Table 5. Reference Node

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Behavior when double clicked</th>
<th>Behavior on following a hyperlink</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>jumping to the End Node of the sub-graph that is associated to the preceded Episode Node</td>
<td>not meaningful</td>
</tr>
</tbody>
</table>

For edges, it is not meaningful to define double click actions or hyperlinks. The edges are not intended to carry topical subject content, but didactics of a (mandatory, conditioned, or recommended) switch between the nodes of the graph. However, the way how the sequence of nodes is specified needs to be defined, in particular in case of alternatives and forks.

As mentioned above, there is a rule to follow an outgoing edge with the same starting color than the target color of the incoming Edge. This serves expressing conditions for leaving a node that refer to the incoming edge. The edge types are explained in Tables 6-9.

### Table 6. Simple Edge

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>defines a unique successor node</td>
</tr>
</tbody>
</table>
4. ESTIMATING SUCCESS CHANCES OF STORYBOARD PATHS

Here, we outline a concept to estimate success chances of curricula, which are composed by students at SIE of TDU in their curriculum planning class in the first semester. Since the storyboard representation enjoys a certain degree of formality, there is an opportunity to apply data mining techniques on storyboarding paths that have been used by students. Furthermore, these paths can be associated with the student’s related success, i.e. his/her final result of the study. Based on these examples, the success chance of intended paths can be estimated as follows. The concept is described in detail in (Boeck, 2007). Furthermore, (Boeck, 2007) contains a prototypical implementation in Prolog, which shows its applicability.

4.1 Construction of a Decision Tree

The storyboard developed for TDU (Dohi et al., 2006a; 2006b) models the opportunities to form curricula. Here, the edges specify prerequisite conditions. The start node of an edge specifies a subject that is a prerequisite of the subject, at which the edge ends.

The construction of the decision tree is based on the paths of former students through the storyboards which model the “space of opportunities”, in which the students took a particular one, which is a path through the storyboard. Each of those paths can be associated with the degree of success, which has been achieved by this student. In case a set of students went the same path, the degree of success can be estimated by the average degree of all students that went this path.

More concretely, this path starts at the start node of the top level storyboard and ends at its end node. For each episode on this path, the related episode is replaced by its sub-graph. This replacement is continued throughout the entire hierarchy of nested graphs. To sum up, such a path consists of (atomic) scenes only (Boeck, 2007).

Each scene of this storyboard application (Dohi et al., 2006a; 2006b) represents a subject students can enroll. Figuratively speaking, the decision tree is constructed on the basis of a “flatten” storyboard. Flatten, in this context, means the graph hierarchy it “flatten down” to just one level with no subgraph.

The decision tree is based on the concept of bundling common starting sequences (Boeck, 2007) of the various paths to a knob of the tree. In (Boeck, 2007) these starting sequences are called “least common denominator”. Of course, all paths went by students start with the start node, which forms the root of the decision tree. Several first elements will result in several sub-trees right below the root. This continues for
each sub-tree accordingly, i.e. if different paths with a common starting sequence from the root until the actual node differ in their next (subsequent) node, related sub-trees will be established.

Each node in this tree, which represents a final node of a path, is followed by a label-node, which contains a list of marks that students received after going this path along with the number of occurrences (student cases for this mark). Additionally, weighted arithmetic average value (called GAM – “Gewichtetes Arithmetisches Mittel”, in (Boeck, 2007) in German) of these marks is represented in this label. The value of GAM serves as an estimation of success chances for future students that plan to go the same path.

4.2 Utilizing the Decision Tree for Path Estimation and Completion

If a student submits a plan for an intended curriculum, which is already represented in the decision tree (as a path from its root to a node that is succeeded by a label node), the prediction is very easily estimated by presenting the content of this label.

In the other case, i.e. if a student submits a curriculum plan, which is not represented in the decision tree so far, the most similar sub-path in the decision tree will be identified. Similarity, in this context, refers to the number of subjects of starting sequences of all represented paths. In other words, those paths in the tree will be identified, which have the longest starting part in common with the submitted curriculum. The last node of this path forms the root of several sub-trees, which represent remaining paths, which are all different from the submitted remaining path. As the success chance estimation, all label nodes of the sub-trees are merged and their common weighted arithmetic average will be provided. To indicate the degree of similarity, the length of the starting sequence divided by the entire length of the submitted path will be presented.

Of course, in such a case, the student may be interested in suggestions to modify the submitted path in a way that the success chance reaches an optimum. Modifying, in this context, means the exchange of the rest-path, which is submitted, but not represented in the decision tree, by the most successful one that is represented in the tree. Here, the “most successful” alternative rest-path is the one with the best weighted arithmetic average value among the paths represented in the sub-trees starting at the last node, which have the tree and the submitted curriculum in common.

Based on this modification suggestion for the rest path along with the similarity degree between the submitted and proposed path, the student can make a decision on whether or not holding on the submitted curriculum or modifying if according to an optimization of the success estimation.

5. CONCLUSION

In contrary to basic level education such as those in primary and secondary schools, academic education at universities is characterized by (1) a large variety of opportunities to compose academic time lines or class schedules and (2) teachers (professors and tutors), which are usually excellent experts in their subject, but do not necessarily have the didactic skills to teach their subject. In particular, at the SIE of TDU, students are required to be more flexible in designing their study according to their needs, wishes, interests, and talents.

However, there are requirements and rules to guarantee a certain level of academic quality. These rules are often complex and difficult to overview. A remarkable number of students have possibilities of failing by violating such regulations. Students need assistance in this jungle of opportunities and limitations.

A basic property of a qualified guidance is adaptability, i.e. a certain dynamics with respect to varying learning needs, context conditions, and the students’ educational history. At a first view, the basic benefit of storyboards compared to any complete representation of rules is its easy overview on relevant class schedules by (1) nesting the graphs and (2) reducing them down to the individually possible choices according to the particular students’ needs.

The deeper benefit is far beyond that. Storyboarding is a step towards making academic education processes a subject of reasoning with AI technologies like Data Mining and finally identifying successful didactic patterns. This is possible due to the fact, that storyboards have a certain degree of a formal knowledge representation, which is (1) controlled by a set of construction operation that ensures formal correctness when designing a storyboard and (2) verified for further formal features by automatic structure tests after designing a storyboard.
This opens the door to design learning plans that a priori ensure a certain degree of learning quality, i.e. a strong indication that learning ends up with a high level of success in academic education as a result of incrementally refining the storyboard based on an automatic analysis such as Data Mining of its usage. In other words, storyboarding supports quality management in academic education.

In particular, we developed a technique that allows an estimation of success chances of a curriculum, which is designed by the students of SIE at TDU be themselves in their Curriculum planning class in the first semester. Moreover, this estimation can be supplemented by a plan modification proposal, which aims at optimizing the success chances. We sketched a concept to perform this feature. To prove the applicability, this concept has been prototypically implemented (Boeck, 2007).

Our upcoming work is directed towards solving the following issues:

(1) Of course, this implementation needs to be practiced at SIE of TDU. Also, some representative data (example paths and related success) needs to be gained.

(2) A definition and representation of (formally to check) criteria, which allow the specification of individual goal-driven storyboards. In fact, this is very different in different cultures, countries, and universities. Therefore, we plan to do that prototypically for the SIE at TDU.

(3) Storyboards have a high performance with respect to didactical issues of planning education processes. However, there is (still) no capability to manage these processes according to their resources (e.g. to concretely planning weekly timetables based on requests and available capacities like rooms, teachers, equipment and so on). Therefore, a desirable synergy effect is expected when incorporating the capabilities of the Dynamic Syllabus tool of the DLNRS into the storyboards.

(4) Also, individual learning plans should not only be based on individual quantitative capability issues (like GPA) or the success of former students, who went similar ways. But also individual properties, talents and preferences should be considered. Some students are more talented for analytical challenges, others are more successful creative or composing tasks and others have an extraordinary talent to memorize a lot of factual knowledge. Consequently, we need to include a user profile to avoid lavishing the students with suggestions that don’t match their individual preferences and talents.

REFERENCES


