

Knauf, Rainer; Böck, Ronald; Sakurai, Yoshitaka; Tsuruta, Setsuo :

A priori evaluation & refinement of curricula by data mining over storyboards

Zuerst erschienen in:

Proceedings of the Twenty-First International Florida Artificial Intelligence Research Society Conference : 15 - 17 May 2008 in Coconut Grove, Florida, USA / ed. by David C. Wilson ... - Menlo Park, Calif. : AAAI Press, 2008. - ISBN 978-1-577-35365-2, S. 335-340

A Priori Evaluation & Refinement of Curricula by Data Mining over Storyboards

Rainer Knauf, Ronald Boeck

Faculty of Computer Science and Automation
Technical University Ilmenau
PO Box 100565
98684 Ilmenau, Germany
rainer.knauf@tu-ilmenau.de

Yoshitaka Sakurai, Setsuo Tsuruta

School of Information Environment
Tokyo Denki University
2-1200 Musai-gakuendai
Inzai, Chiba, 270-1382, Japan
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. 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 storyboarding. By using data mining-like methods to evaluate storyboard paths, we are able to estimate success chances of storyboard paths. Based on this 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.

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 could not comply with some of these regulations or did not 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.

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 and 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 (Flehsig 1996), and storyboards may provide the expressive power suitable to the design and implementation of learning processes.

Storyboarding as introduced in (Jantke and 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, data mining can be performed over the paths of particular students after they completed their study at TDU. This will finally make the students able to create curricula with an optimal chance of success.

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 nodes. The tree's leaves are the (known) success level (examination results).

The estimation of success chances for a newly built schedule is performed through 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.

Dynamic learning need reflection – A concept that includes individually compose curricula

In its concept, DLNRS (Dohi and 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. The features of DLNRS, which opened the door for storyboarding the search space for curricula, are (1) the abolition of the traditional rigid academic year with fixed curricula, and (2) the abolition of compulsory subjects. Other key features of this concept are (3) the replacement of a fixed charge per year by a subject-oriented paying system, (4) a change in the class period length, and (5) the introduction of so called Grade Points and the Grade Point Average (GPA).

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.

Specific compulsory subjects have been replaced by the concept of prerequisite conditions. These conditions are expressed in two levels of recommendation, namely (a) sub-

jects that have to be learnt before and (b) 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.

The introduction of the DLNRS at SIE is supported by a Curriculum Planning Class, which aims at developing an individual curriculum for each student by himself that meets his needs and desires.

The DLNRS was introduced in 2004. The effects of the DLNRS were investigated through a questionnaire at the end of the Curriculum Planning course. 203 students were asked. The rate of understanding the prerequisite conditions was about 60 %, i.e. the method of displaying the prerequisite condition needs to be improved (by storyboarding). The rate of useful class schedules for a current semester and until graduation was more than 80 %. Thus, the DLNRS and the tool to compose curricula seems to be effective and is going to become more effective, if the prerequisite condition display method is improved. The significantly decreased time frame in which students create their timetable supports this thesis about the DLNRS. More than 50 % of the students were able to create their class schedules in two hours or less. About 50 % of the students were not able to complete their class schedule in the Curriculum Planning class. Those students were given a week to complete it as homework. A total of 162 out of 173 (96.3 %) students were finally able to send their class schedules.

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.

Storyboarding – A concept to model the search space for curricula composition

The storyboard approach adopted here (Jantke and Knauf 2005; Knauf and 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, a current situation (like being under time pressure), and other circumstances. The various alternative paths to study this paper, for example, may be driven by the reader's role as follows:

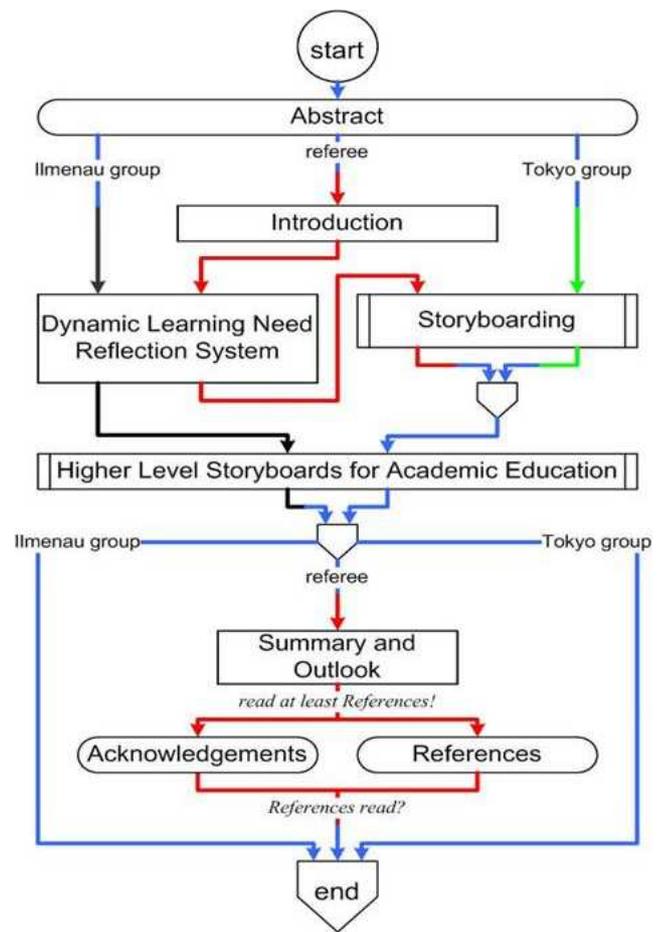


Figure 1: An exemplary storyboard

- 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 also 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, on the other hand, 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 repre-

sentation 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.

Storyboards can be exploited for representing the didactics or tutorial knowledge of academic education, especially of DLNR. Through this, learning plans have several advantages. These plans are such as students' curriculum plans for getting their academic career. Besides being visual in nature, they are formally created and can also be dynamically verified and modified each periodical usage or each semester.

Namely, in order to exploit the storyboard concept for an academic graduation career, a new method of storyboard usage or high level dynamic usage of storyboards, that we call "dynamic storyboarding", is proposed. The results of a feasibility study are shown here.

In dynamic storyboarding, subjects to be taught are represented as scenes. This is opposed to the original usage of storyboards where subjects are represented as top level episodes. As a matter of fact, in learning plans or curriculum plans, some higher level structure (i.e. episodes) lies above the subject level.

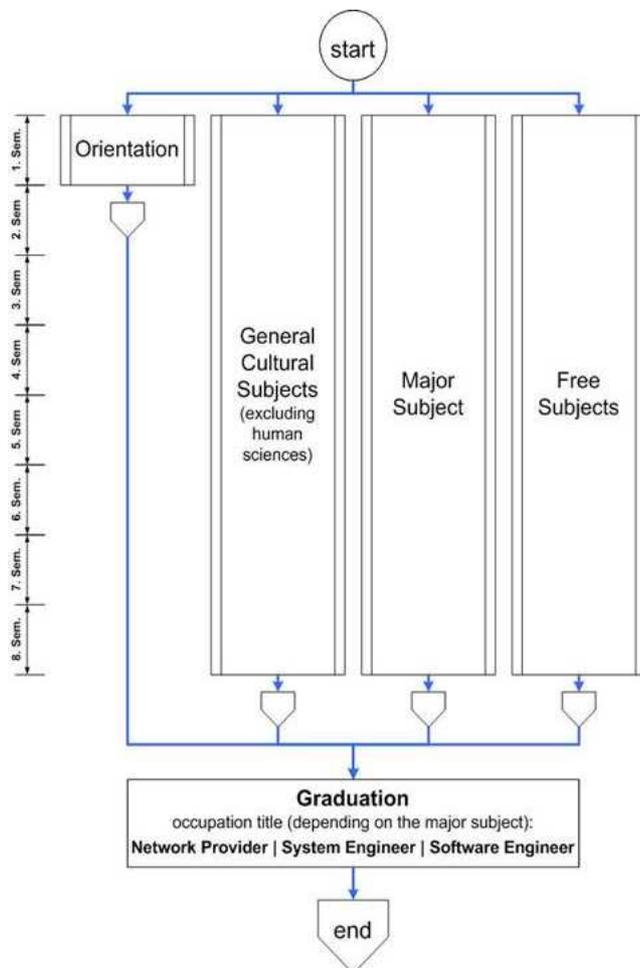


Figure 2: Top level storyboard for graduation

As shown in Figure 2, "dynamic storyboarding" provides a chance to make explicit the higher level complex and dynamic education structure such as for a university study. This makes such complex and dynamic education structure (or its overview) more understandable to students.

A priori path evaluation – A concept to estimate 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.

Construction of a decision tree

The storyboard developed for TDU (Dohi et al. 2006a; Dohi et al. 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 the student. In case a set of students went the same path, the degree of success can be estimated by a weighted average degree of them.

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.

Each scene of this storyboard application (Dohi et al. 2006a; Dohi et al. 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 is "flattened down" to just one level with no subgraph.

The implementation is realized in Prolog. The input is implicitly given by a database of predicates describing the storyboard. The storyboard is represented by Prolog facts

- `includes(graph, [<elements>])`
- `edge([<begin>], [<end>], [<(color, color)>])`

The edge-fact models the aspect of transition between nodes. They are, however, not used for decomposing. The method is based on the includes-predicates only and is performed as follows:

- First, we distinguish sequential from parallel structures. Sequential elements form a sub-path, parallel structures form a single element that is handled as one element of a path.
- Because the input is a list, we check each element, whether it is an episode by using the predicate includes.
- In case it is an episode, the elements of its sub-graph will be decomposed recursively.
- All atomic elements will be appended to the resulting path, which is the output of the method.

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 (GAM) 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.

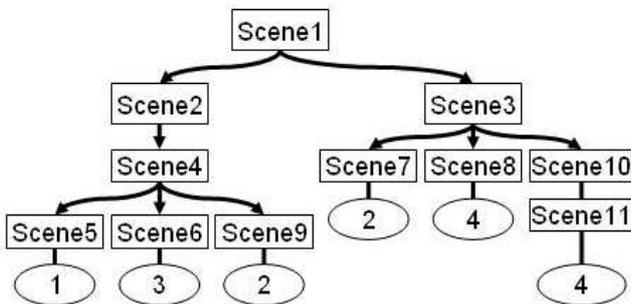


Figure 3: A decision tree

Figure 3 shows a small example of a storyboard-tree. For simplicity, the labels (elliptic nodes) are reduced to the value of GAM. According to the form of the scenes in a storyboard the atomic attributes of the tree are placed in rectangles. The path through the storyboard-tree is defined by its directed edges. Only the connection to the label is non-directed, because it refers to the complete path. In (Boeck 2007) a prototypical implementation of the procedure is given. The decision tree to be built is represented by Prolog facts

- `tree(root, [<subtrees>])`

and the success labels that form the leaves of the tree are represented by Prolog facts

- `assess(node, <statistics>)`

The statistics-parameter of the assess-predicate are the values which are defined in the label.

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 it according to an optimization of the success estimation.

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 regu-

lations. 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 by 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 checked) 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. For example, some students are more talented for analytical challenges, others are more successful creative or composing tasks and others may have an extraordinary talent to memorize a lot of factual knowledge.

Consequently, at some point we need to include some sort of user profile to avoid lavishing the students with suggestions that don't match their individual preferences and talents.

References

- Boeck, R.: *Ein Data Mining Verfahren zur Pfadbewertung in Storyboards*. (German), Diploma Thesis, Technical University Ilmenau, Faculty of Computer Science and Automation, 2007.
- Dohi, S. and S. Nakamura, S.: The development of the Dynamic Syllabus for School of Information Environment. *ITHET03*, pp. 505-510, 2003.
- Dohi, S., Nakamura, S., Sakurai, Y., Tsuruta, S., and Knauf, R.: Dynamic Learning Need Reflection System for academic education and its applicability to Intelligent Agents. *Proc. of the 6th IEEE International Conference on Advanced Learning Technologies (ICALT 2006)*, Kerkrade, the Netherlands, Los Alamitos, CA: IEEE Computer Society, ISBN 0-7695-2632-2, pp. 459-463, 2006.
- Dohi, S., Sakurai, Y., Tsuruta, S., and Knauf, R.: Managing academic education through dynamic storyboarding. *World Conference on e-Learning in Corporate, Government, Healthcare, and Higher Education 2006 (E-Learn 2006)*, Honolulu, Hawaii, USA, Chesapeake, VA: Association for the Advancement of Computing in Education (AACE), ISBN: 1-880094-60-6, pp. 1611-1619, 2006.
- Flechsig, K.-H.: *Kleines Handbuch didaktischer Modelle*. (German), Neuland, 1996.
- Jantke, K.P. and Knauf, R.: Didactic design through storyboarding: Standard concepts for standard tools. *Proc. 4th Internat. Symposium on Information and Communication Technologies, Workshop on Dissemination of e-Learning Technologies and Applications (DELTA 2005)*, Cape Town, South Africa, 2005, N.Y.: ACM Press, ISBN 0-9544145-6-X, pp. 20-25, 2005.
- Knauf, R. and Jantke, K.P.: Storyboarding - An AI technology to represent, process, evaluate, and refine didactic knowledge. Jantke & Kreuzberger (eds.): *Knowledge Media Technologies. Proc. of the First Core-to-Core Workshop*, Dagstuhl Castle, Germany, 2006, TU Ilmenau, ISSN 1617-9048, pp. 170-179, 2006.
- Knauf, R., Sakurai, Y., and Tsuruta, S.: Toward Making Didactics a Subject of Knowledge Engineering. *The 7th IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, Niigata (Japan), Los Alamitos, CA: IEEE Computer Society, ISBN 0-7695-2916-X, pp. 788-792, 2007.