Efficient Exam Correction at Scale

Streamlining Paper-Based Assessments with the VoLI-KOrN System

Jonas Betzendahl, Dominic Lohr, Marc Berges, Michael Kohlhase

Abstract: *Dieser Beitrag stellt VoLl-KOrN vor, ein System zur Digitalisierung und Optimierung von papierbasierten Prüfungen an Hochschulen. Durch QR-Codes auf den Prüfungsunterlagen ermöglicht VoLl-KOrN die automatische Zuordnung gescannter Prüfungsunterlagen zu Studierenden. Über eine Weboberfläche können Lehrende die Prüfungsaufgaben einsehen, organisieren und bewerten. Die Innovation liegt hierbei in der Verwendung des Konzeptes der Antwortklassen, welche ähnliche Antworten gruppieren und so eine Bewertung mit Fokus auf Objektivität und Konsistenz ermöglicht. Darüber hinaus ermöglicht VoLl-KOrN eine systematische Analyse der Prüfungsdaten zur Verbesserung der Lehrstrategien und zur Ermittlung von Lernlücken. Dies eröffnet weitere Möglichkeiten wie die Anbindung an Lernplattformen, um z. B. ein Lernendenmodell auf Grundlage der Leistungen zu aktualisieren.*

This paper introduces VoLl-KOrN, a tool designed to digitize and optimize paper-based examination processes at universities. By utilizing QR codes on exam papers, VoLl-KOrN allows for submissions to be scanned and automatically assigned to students. The web interface allows educators to view, organize, and assess scanned exams. Its innovation lies in its use of answer classes that cluster similar responses, allowing automatic or manual scoring with a focus on objectivity and consistency. Additionally, the system enables a systematic analysis of exam data to improve teaching strategies and identify learning gaps. This enables further opportunities like interconnecting with learning platforms to, e.g., update a learner model based on exam performance.

Keywords: *Klausurkorrektur; Antwortklassen; Lerndiagnose; Prüfungsmanagement; Digitale Lehre; Lernendenmodell; Assessment; Answer classes; learning analytics; educational technology; digital teaching; learner model*

1. Introduction

The grading of large-scale exams in higher education is a challenging task regarding the organization of the grading process (e.g., finding suitable graders, synchronizing people grading the same task), as well as providing valuable feedback to the students.

Addressing those problems, we introduce VoLl-KOrN, a web application designed to digitize and optimize paper-based examination processes at universities. The system allows educators to process exams more efficiently, ensures that feedback can be provided more quickly and adaptively, and provides valuable insights into teaching processes. Once the scanned exams are available in the system, they can be selectively searched and assigned to reviewers without passing around stacks of paper. The assignments can be viewed, organized, and evaluated by graders via a web interface – accessible from anywhere unless restricted to a local network. However, the real innovation of VoLl-KOrN lies in the grading process itself using the concept of *answer classes* – specifications of answer clusters that are defined by objectively observable criteria (see Section 2). Foremost, this leads to more objective and consistent grading and improves the quality of feedback for students, as answer classes may include general comments and enable the inclusion of a learner model to provide learner-specific feedback.

Additionally, VoLl-KOrN facilitates a systematic analysis of exam outcomes. Data relating to the distribution of submissions across answer classes offers valuable insights into the teaching and learning processes. Instructors can easily discern the subjects showing notable weaknesses, signaling areas where the educational focus may have been insufficient, as well as pinpoint disparities in student responses warranting particular attention.

Another advantage is the possibility of interconnecting with learning platforms. Since modern adaptive learning systems usually contain a learner model in which information about the individual student's learning process is stored, VoLl-KOrN offers the possibility, e.g., for updating the models of the exam participants based on their performance. As a result of this compatibility, personalized learning recommendations can be made to better support students' individual needs.

2. The Concept of Answer Classes

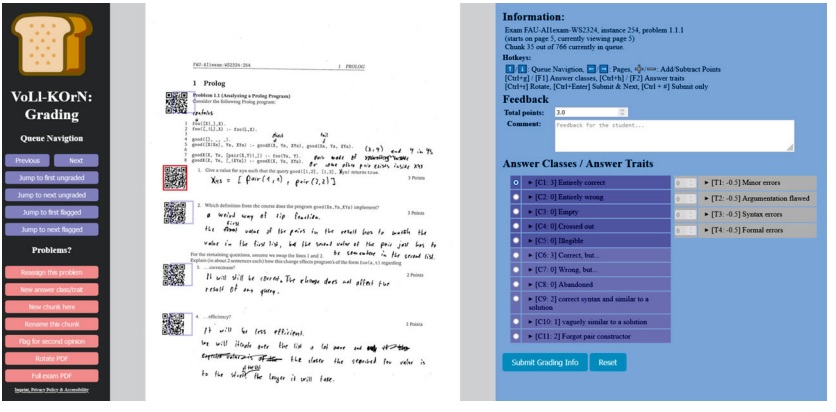
The concept of *answer classes* (AC) is part of the Y-model framework for formalizing computer science tasks in the context of adaptive learning systems (Lohr, Berges, Kohlhase, Müller & Rapp, 2023). They can be understood as a set-theoretic propositional form where answers are assigned to an answer class if they meet the criteria of a class description. For a reliable assignment to ACs, this description must be (1) un-

ambiguous (free of interpretation) and (2) objectively observable. This ensures that educators arrive at the same classification of an answer independently of each other and that a task's state becomes more comprehensible for both educators and learners. A distinction is made between ACs that are specific to a task – and therefore only exist in the context of a task or task type – and ACs that are valid across tasks. When developing ACs, care is taken to ensure that they are didactically meaningful and, at best, reusable. For a more detailed overview of the concept, see Lohr, Berges, Kohlhase & Rabe, 2023.

3. The VoLI-KOrN System

At the outset, the exams are scanned and uploaded to the VoLl-KOrN system together with a file containing relevant meta-information like task identifier, maximum number of points, and a list of available ACs. As we use semantic annotations in the learning objects underlying our adaptive learning assistant, this information can automatically be generated from the source code with the same plugin that autogenerates the QR codes. Since we are working on scans of actual paper exams, these QR codes – printed in the margins next to every (sub-)problem and on every page of the exam – make it possible to identify students' answers at any time. It no longer matters in which order the exams are scanned or whether they were mixed up after being stapled apart.

Figure 1: Screenshot of the VoLl-KOrN grading interface



The QR codes encode the following information: (1) an **Exam Identifier** that is unique across courses, semesters, or potential retakes (e.g., `UNI:AI2-WiSe23-`

`retake`), (2) a **Content Identifier** (`e.g., problem-2.1.5`) and (3) a **Student Identifier** to match scans to individual students.

Digitizing the grading process allows graders to work from anywhere at the same time – provided they have a working internet connection. VoLl-KOrN implements standard web security mechanisms for authentication and authorization to prevent misuse.

3.1 Grading Workflows in VoLl-KOrN

The first step is to assign problems to graders, which results in a grading queue for each grader. For this, instructors (or graders themselves) can search the QR-code identifiers via regular expression and limit the search to content with certain tags (e.g., *not yet graded*, *flagged for review*).

Afterward, graders go through their assigned queue step by step. The web interface (see Figure 1) has three panes: one on the left for queue navigation and non-grading interactions (such as creating new ACs or rotating the page to counteract mistakes during scanning), another (in the center) that shows the page that contains the current content (with the relevant QR-code highlighted in red), and a third one on the right for grading interactions.

To evaluate a (sub)task, educators select ACs that match the student's answer. Every answer class comes with point recommendations for scoring and learner feedback specific to the answer class. Both can manually be overwritten and adapted by the grader if necessary. While working with VoLl-KOrN it quickly became clear that the theoretical concept of ACs from [1] needed to be refined to be practical. So VoLl-KOrN provides a set of 'standard' answer classes that apply to all (sub)-problems, e.g., *totally correct*, *empty*, or *illegible*, and allows graders to create new ones by need. In addition, the system adds general modifiers like '*minor errors*', '*argumentation flawed*', and answer-class specific ones like '*but (correctly) mentioned …*' that can be combined with an answer class. The modifiers come with an adjustment of the score. In our experience with almost 2000 graded exams, this combination method allows us to adequately and efficiently use the – rather sizable – set of answer classes without overloading graders.

It is always possible to export a current state of grading for any given exam to be imported into, for instance, a spreadsheet processor of choice to be merged with additional data (such as students' points from homework exercises) or extended by calculations (such as resulting grade, …). This export can also contain information on which answer classes were selected for which problem and how many times (and even on which instance, which can serve as input to compatible adaptive learning platforms to update the student's learner model).

VoLl-KOrN allows the re-export of the answer classes recorded in a grading session into the problem sources so that they can be shared (and adapted) for later exams and homework assignments.

Lastly, VoLl-KOrN offers functionality to give students feedback on their work, for example, during exam review. For any specific instance (read: student), it can create a feedback digest that lists all problems of the exam with the number of (possible) points as well as the selected answer classes and any grader comments. This provides learners a useful overview of what they did correctly or wrongly that is more accessible and approachable than having to file through the exam itself and trying to decipher usually curt and handwritten notes from graders.

4. Conclusion, Limitations, and Future Work

The VoLl-KOrN system currently adapts the paper-based assessment workflow mandated in our university, but it would be easy to adapt it to online- testing approaches where exams are taken digitally. Additionally, the system is geared towards human raters to assign students' answers to answer classes. However, we have already made efforts to build automatic classifiers to incorporate them into VoLl-KOrN in the future. Finally, the grading workflow based on answer classes is sufficiently systematic that it can be used for peer-grading workflows. First experiments show that the added reliability and consistency enable feasible peer grading even with peers less knowledgeable than usual graders, like instructors and teaching assistants. Indeed, there is an anecdotal indication that being exposed to answer classes furthers the understanding of grading peers. We plan to integrate this into an adaptive learning assistant in the near future.

Our experience has shown that the answer classes collected during a grading session are a very valuable educational resource that can provide instructors and instructional content developers with crucial information about the efficacy of learning materials and approaches. We want to develop this further towards truly data-driven educational practices.

Overall, VoLl-KOrN offers a comprehensive solution for improving the assessment of exams at universities, especially with many participants. It increases the quality of feedback and enables data-driven improvement of teaching. In so doing, VoLl-KOrN promises to increase the efficiency and effectiveness of university teaching.

Literaturverzeichnis

Lohr, D., Berges, M., Kohlhase, M., Müller, D., & Rapp, M. (2023). The Y-Model – Formalization of Computer Science Tasks in the Context of Adaptive Learning Systems. *2023 IEEE 2nd German Education Conference (GECon)*. IEEE. http://doi.org/10.1109/GECon58119.2023.10295148

Lohr, D., Berges, M., Kohlhase, M., & Rabe, F. (2023). The Potential of Answer Classes in Large-scale Written Computer Science Exams. In J. Desel & S. Opel (Hrsg.), *Hochschuldidaktik Informatik (HDI) 2023* (S. 179–189). GI.