# CITUC: AUTOMATISIERTE LÖSUNGSBEWERTUNG IM E-LEARNING DURCH KOLLABORATIVES FILTERN

Frank Loll, Niels Pinkwart[1]

***Kurzfassung***
*Kollaborative Filtertechniken spielen im heutigen Web 2.0 eine Schlüsselrolle. Während ihr Einsatzgebiet bislang überwiegend im E-Commerce Bereich zu finden ist, besitzen sie auch Einsatzpotenzial im E-Learning. Das in diesem Artikel beschriebene CITUC-System kombiniert kollaboratives Filtern mit Peer Reviews, um Systemfeedback zu studentischen Aufgabenlösungen zu geben. Das webbasierte System wurde mit Erfolg begleitend zum Übungsbetrieb bei einer Vorlesung an der Technischen Universität Clausthal eingesetzt und übertrug damit die Erwartungen aus einer vorherigen Laborstudie in die Praxis.*

## 1. Einleitung

Kollaborative Filteralgorithmen [4] bilden eine wesentliche Basis des „Web 2.0“. Diese Familie von Algorithmen ist dadurch gekennzeichnet, dass Assoziationen zwischen Benutzern und System-Artefakten durch explizite oder implizite Benutzeraktionen ermittelt und für Zwecke des Systems genutzt werden. Beispiele für diese Assoziationen sind Buchbestellungen bei amazon.com, die Eingabe von Profilen bei Online-Partnerschaftsvermittlungen und das Tagging von Bildern bei flickr.com. In diesen Anwendungen werden die gespeicherten Assoziationen zur Empfehlung von Artefakten (Büchern, Personen, Bildern, etc.) genutzt. Die Details der Berechnungsverfahren unterscheiden sich dabei zwischen den einzelnen Systemen, das zugrunde liegende Prinzip des kollaborativen Filterns – die Ausnutzung von Benutzerinformationen zur Bewertung und Empfehlung von Artefakten im System – ist jedoch ein gemeinsamer Aspekt vieler Web 2.0-Anwendungen.

Auch in E-Learning-Systemen haben kollaborative Filteralgorithmen deutliches Anwendungspotenzial: Die Qualität einer Aufgabenbearbeitung durch Studenten kann mittels dieser Verfahren durch Bewertungen anderer Studenten heuristisch ermittelt werden (Peer Review). Das Ziel des kollaborativen Filterns besteht in diesem Fall weniger in der Berechnung einer potenziellen Passung zwischen Benutzern und Artefakten (wie in den klassischen Anwendungsbereichen), sondern in der Schätzung der Qualität von Lösungen. Ein solcher Ansatz hat viele praktische Vorteile. Er entlastet Lehrkräfte und Tutoren und ermöglicht es gleichzeitig Studenten, ihre

[1] Technische Universität Clausthal, Institut für Informatik

Evaluations- und Kritikfähigkeiten zu trainieren, indem sie Aufgabenbearbeitungen anderer Studenten bewerten. Falls Aufgaben offener gestellt sind und mehrere verschiedene Lösungen zulassen, können die Lernenden durch die Analyse und Bewertung von Lösungen anderer Lernender auch alternative Lösungsansätze bzw. Sichtweisen kennen lernen.

Typische Kritikpunkte eines solchen Peer Review-Ansatzes in der Lehre bestehen darin, dass Lernende aufgrund mangelnden Wissens oder Erfahrung Fehler beim Bewerten von Lösungen machen können. Auch läuft ein solches Verfahren Gefahr, absichtlich manipuliert zu werden [2, 9]. Jedoch haben kollaborative Filteralgorithmen als Basis eines Peer Reviews im E-Learning auch Vorteile gegenüber klassischen Ansätzen, in denen die Lehrkraft (bzw. erfahrene Assistenten) Lösungen korrigieren: Lernende können sich oft in Probleme anderer Lernender gut einfühlen und verstehen daher die Gründe für fehlerhafte Aufgabenbearbeitungen manchmal besser als Experten [5]. Speziell in unstrukturierten Domänen [7], in denen Aufgaben typischerweise keine einfache Richtig/Falsch-Klassifikation von Lösungen zulassen, sind Multiple-Choice-Tests unangemessen und offenere Aufgabenstellungen die Norm. Die Bewertung von Aufgabenlösungen ist in diesen Gebieten zeitaufwändig, schwierig und bislang nicht automatisierbar. Ein E-Learning-System kann es auf Basis kollaborativer Filteralgorithmen erlauben, dass Lernende gegenseitig ihre Lösungen charakterisieren und annotieren. Zusammengenommen können diese vielen Meinungen anderer Lernender sogar qualitativ besser sein als die Meinung eines einzelnen Experten [1]. Dieses Phänomen wird allgemein als „Kollektive Intelligenz“ oder „Wisdom of the Crowds“ [11] bezeichnet.

## 2. Kollaboratives Filtern im E-Learning

Trotz ihres Potenzials sind kollaborative Filtermechanismen im E-Learning-Sektor bisher selten eingesetzt worden und es wurden bislang nur relativ wenig empirische Studien über die Effektivität dieser Verfahren publiziert. Eines der existierenden Systeme ist Peer Grader (PG) [3]. Dieses System erlaubt ein Peer Review von Aufgabenlösungen (Studenten bearbeiten erst Aufgaben selbst und können dann Lösungen anderer Studenten zu diesen Aufgaben bewerten) und ermöglicht es zusätzlich, auch die Bewertungen selbst zu bewerten (über die Frage: „Wie hilfreich war das Review für die Überarbeitung der Lösung?“), um qualitativ hochwertiges Feedback zu motivieren. Dieser mehrstufige, komplexe Begutachtungsprozess von PG ist jedoch zeitintensiv, was in der Praxisnutzung zu Problemen führen kann [3].

Ein neueres und auf literarische Anwendungen spezialisiertes System ist SWoRD (Scaffolded Writing and Rewriting in the Discipline) [1]. Dieses System benutzt den (ebenfalls zeitaufwändigen) Prozess der Begutachtung von Zeitschriftenmanuskripten und wendet diesen im E-Learning an. Evaluationen zu SWoRD zeigten die Eignung des Systems und belegten insbesondere, dass mehrere Bewertungen von Studenten hilfreicher sein können als einzelne Bewertungen durch Experten [1].

Im LARGO System [10] wird kollaboratives Filtern dazu verwendet, um Argumentationsfertigkeiten bei Jurastudenten auszubilden. Lernende erstellen in diesem System grafische Repräsentationen von gerichtlichen Argumentationslinien und können Elemente der Diagramme von anderen Lernenden bewerten. Durch einen kollaborativen Filteransatz gelangt LARGO zu Wissen über die Qualität von Argumenten in den Diagrammen.

Ausgehend von den Erkenntnissen, die wir im Rahmen der Untersuchung dieser bestehenden E-Learning-Systeme gewonnen haben, bestand unser Bestreben darin, ein eigenes System zu ent-

wickeln, welches die Schwächen der bestehenden Systeme, d.h. den zeitraubenden Ablauf (SWoRD, PG) sowie die Domänenabhängigkeit (LARGO, SWoRD) eliminiert und gleichzeitig die jeweiligen Stärken, d.h. die geringe Anzahl benötigter Reviews (SWoRD) und die sofortige Rückmeldung (LARGO) erhält. Unser Ansatz sowie dessen Umsetzung und Evaluation im Praxiseinsatz werden in den folgenden Abschnitten des Artikels beschrieben.

## 3. Vorhergehende Untersuchungen

Als Vorarbeit [6] entwickelten wir eine erste Heuristik, welche die Techniken des Peer Reviews (PG, SWoRD) sowie des kollaborativen Filterns (LARGO) kombiniert. In einer Laborstudie bekamen 45 Teilnehmer die gleichen Fragen gestellt, die sie selbstständig lösen sollten, ehe sie alternative Antwortvorschläge von anderen Teilnehmern zur jeweiligen Frage in anonymisierter Form bekamen. Diese alternativen Lösungen wurden auf einer Skala von 0 (sehr schlecht) bis 10 (sehr gut) bzw. mittels einer Gut-/Schlechtklassifikation bewertet. Abschließend schätzten die Teilnehmer ihre eigenen Leistungen entsprechend ein. Die Reihenfolge der Fragen war festlegt und es konnte keine Frage übersprungen werden. Die kollaborative Filterheuristik ermittelte, basierend auf den gegenseitigen Bewertungen, eine Qualitätseinstufung für jede Lösung, die anschließend mit einer Experteneinstufung verglichen wurde. Hierbei konnten wir zeigen, dass die Bewertungen der Heuristik, unter streng kontrollierten Bedingungen, bei vier oder mehr Bewertungen einer Lösung eine hohe Übereinstimmung mit Expertenbewertungen aufwies und die Selbsteinschätzung der Teilnehmer signifikant übertraf. Die Heuristik zeigte jedoch Schwächen im Erreichen von extremen Bewertungen. Dieser Punkt wird in den Abschnitten 4.3 und 6.1 wieder aufgegriffen.

## 4. Das CITUC-System

### 4.1. Forschungsfragen

Der nächste Schritt unserer Forschungen bestand nun darin, die Erkenntnisse aus dem Labor in die Praxis zu übertragen. Hierzu verfeinerten wir die Heuristik, entwickelten ein webbasiertes E-Learning-System namens CITUC (basierend auf PHP und mySQL), welches im folgenden Abschnitt beschrieben wird, und setzten dies als begleitendes System zur Klausurvorbereitung (als Ersatz für ein letztes Tutorium in einer universitären Lehrveranstaltung) ein. Hierbei stand die Frage, ob das System das Potenzial hat, klausurvorbereitende Präsenz-Übungen zu ersetzen, im Vordergrund. Weiterhin sollte geklärt werden, ob eine freiwillige Nutzung zu einer aktiven Nutzung des Systems führt (Quantität) und ob das System durch Übungsleiter und Studenten als hilfreich angesehen wird (Qualität).

### 4.2. Systembeschreibung

Nach dem Einloggen in das System mittels anonymen Kennungen bietet CITUC zuerst einen personalisierten Überblick über die Änderungen seit dem letzten Login des jeweiligen Nutzers (s. Abb. 1) und gleichzeitig eine Auswahl der mögliche Aktionen. Im Vordergrund des Systems steht die Aufgabenbearbeitung. Hierbei werden alle vom Dozenten oder von anderen Studenten (s.u.) eingestellten Aufgaben aufgelistet, so dass sich jeder Student Aufgaben heraussuchen kann, die er bearbeiten möchte. Nach Eingabe einer Lösung (siehe Abb. 2) werden dem Studenten alternative Lösungen von anderen Studenten zu der eben bearbeiteten Aufgabe anonym präsentiert. Bei der Zusammenstellung wird die Anzahl bisher abgegebener Bewertungen berücksichtigt und jene Lösungen bevorzugt, die bislang die wenigsten Bewertungen erhalten haben. Diese müssen

anschließend auf einer Skala von 0 (sehr schlecht) bis 10 (sehr gut)[2] bewertet und (optional) kommentiert werden, um den jeweiligen Autoren Hilfe bei der Identifikation von Fehlern zu geben (siehe Abb. 3). Zu jeder bearbeiteten Aufgabe können die Studenten alle bisher abgegebenen alternativen Lösungen und deren Kommentare einsehen. Zusätzlich dazu wird jeder Lösung eine Bewertung vom System zugeordnet, die sich aus den Einzelbewertungen der Studenten errechnet (siehe 4.3).

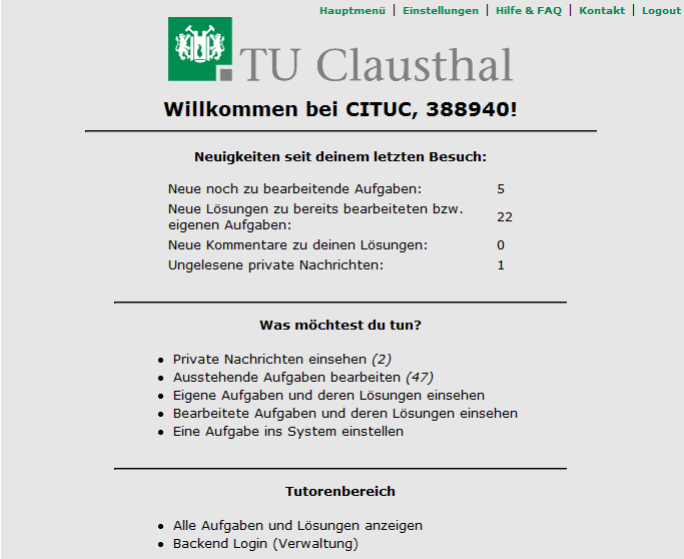


**Abbildung 1: CITUC-Interface mit Awarenessinformationen**

Zur einfachen visuellen Informationsdarstellung wird in CITUC ein 5-stufiges Ampelsystem (sehr schlecht, schlecht, mittelmäßig, gut, sehr gut) anstatt der errechneten numerischen Lösungsbewertungen verwendet. An dieser Stelle können Nutzer durch private Nachrichten oder weitere Kommentare Rückfragen zu den Kommentaren stellen.

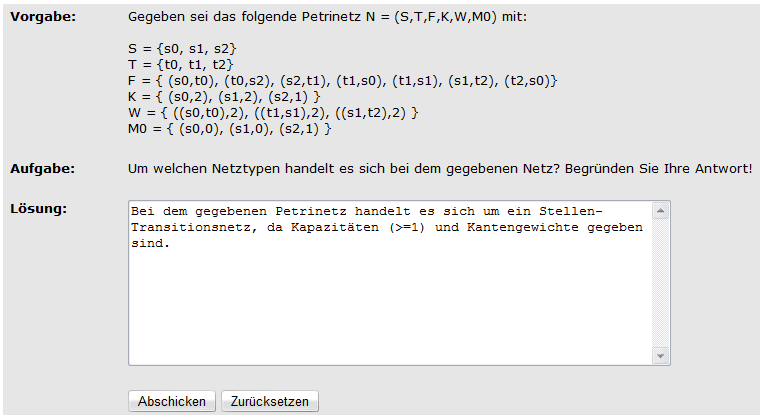


**Abbildung 2: Aufgabenbearbeitung**

Zusätzlich bietet das System den Studenten die Möglichkeit, auch selbst Fragen einzustellen. Im Gegensatz zu den Studenten können die Tutoren bzw. der Dozent *jederzeit alle* Aufgaben und deren Lösungen einsehen (siehe Tutorenbereich in Abb. 1). Dadurch ist es möglich, bereits einge-

[2] Systemintern wurde mit Werten zwischen 0 und 1 gearbeitet, die dann entsprechend mit dem Faktor 10 multipliziert wurden.

tragene studentische Lösungen zu kommentieren, um die automatisierte Bewertung des Systems durch Verbesserungsvorschläge zu unterstützen. Da es sich um ein asynchrones System handelt (d.h. es ist keine zeitgleiche Nutzung des Systems durch Studenten notwendig), dessen Nutzung über mehrere Wochen in der Klausurvorbereitungsphase geplant war, wurden die Teilnehmer zusätzlich täglich bzw. auf Wunsch wöchentlich durch E-Mails mit Awarenessinformationen (neue Aufgaben, neue Lösungen zu bereits bearbeiteten Aufgaben, neue Kommentare zu eigenen Lösungen) versorgt.

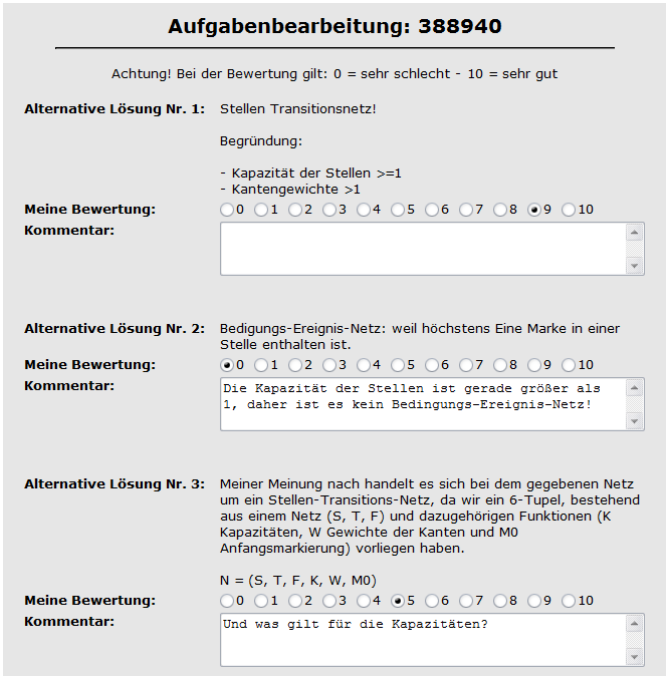


**Abbildung 3: Bewertung alternativer Lösungen mit Kommentarfunktion**

## 4.3. Kollaborativer Filteraspekt der eingesetzten Heuristik

Die von uns eingesetzte Heuristik dient zur automatisierten numerischen Bewertung von abgegebenen Lösungen zu einer Aufgabe auf einer Skala von 0 (sehr schlecht) bis 1 (sehr gut). Diese Heuristik, die auf dem Konzept des kollaborativen Filterns beruht, besteht aus zwei Komponenten.

Die erste Komponente basiert auf der Annahme, dass ein Teilnehmer, der gegebene Lösungsvorschläge auf einer gegebenen Skala korrekt klassifizieren kann, selbst in der Lage ist, eine qualitativ hochwertige Lösung für die jeweilige Aufgabe zu liefern. Darauf aufbauend wird im ersten Schritt der *Basiswert* seiner eigenen Lösung ermittelt. Wenn ein Teilnehmer $n$ Bewertungen $w_1, \ldots, w_n$ für $n$ Lösungen anderer Teilnehmer jeweils mit einem *Qualitätwert* (d.h. mit aktueller Heuristikbewertung) von $q_1, \ldots, q_n$ abgegeben hat, dann berechnet sich der *Basiswert* $b$ der eigenen Lösung des Teilnehmers als:

$$b = 1 - \frac{1}{n} \sum_{i=1}^{n} \frac{(|\, w_i - q_i \,|)}{\max(q_i, 1 - q_i)}$$

Hierbei gilt es zu beachten, dass Bewertungen in Schritten von 0,1 möglich waren. Zur Erläuterung sollen die Bewertungen, die in Abb. 3 erkennbar sind, als Beispiel dienen. Der Teilnehmer bewertet die Lösungen $L_i$ mit $(w_{L1} = 0{,}9)$, $(w_{L2} = 0{,}0)$ und $(w_{L3} = 0{,}5)$. Die *Qualitätswerte* (s.u.) der

Lösungen betragen *($q_{L1}$ = 1,0)*, *($q_{L2}$ = 0,2)* und *($q_{L3}$ = 0,5).* Selbstverständlich sind diese Werte dem Teilnehmer nicht bekannt. Dann ergibt sich folglich als *Basiswert b*:

$$b = 1 - \frac{1}{3}\sum_{i=1}^{3}\frac{(|\, w_i - q_i \,|)}{\max(q_i, 1-q_i)} = 1 - \frac{1}{3}\left(\frac{|\, 0{,}9 - 1{,}0 \,|}{1{,}0} + \frac{|\, 0{,}0 - 0{,}2 \,|}{0{,}8} + \frac{|\, 0{,}5 - 0{,}5 \,|}{0{,}5}\right) \approx 0{,}88$$

Der resultierende *Basiswert* von *0,88* ist verhältnismäßig hoch, was dadurch begründet ist, dass der Teilnehmer die gegebenen Lösungen der anderen Teilnehmer relativ präzise bewertet hat. Unter Berücksichtigung unserer Einstiegshypothese war der Teilnehmer folglich in der Lage, selbst eine qualitativ hochwertige Lösung zu liefern. Durch den *Basiswert* und das Bereitstellen von Lösungen mit unterschiedlicher Qualität zum Start des Systems kann das „cold-start"-Problem kollaborativer Filteralgorithmen [8], das beim Einfügen einer neuen Lösung auftritt, gemildert werden. Um zu testen, ob die Heuristik auch ohne Vorgaben zu sinnvollen Ergebnissen führt, wurde bei einigen Aufgaben auf Vorgaben verzichtet und der ersten im System eingetragenen Lösung ein *Basiswert* von *0,5* zugewiesen. Die Basiswertkomponente wurde gegenüber der vorhergehenden Laborstudie [6] modifiziert, da es in der ersten Version Probleme mit dem Erreichen von sehr hohen bzw. sehr niedrigen Basiswerten gab. Diese wurden durch die hier präsentierte Version gelöst (siehe 6.1).

Die zweite Komponente der Heuristik bildet die *Evaluierungsbewertung*. Nachdem der Teilnehmer seine Lösung abgegeben hat, wird diese anderen Teilnehmern zur Bewertung vorgelegt. Die Bewertungen werden gesammelt und gemittelt, wobei eine Gewichtung der Einzelbewertungen erfolgt, in der die Bewertungen (gemäß Annahme) schlechterer Teilnehmer ein geringeres Gewicht erhalten. Die *Evaluierungsbewertung e* berechnet sich als:

$$e = \frac{1}{\sum_{i=1}^{j} q_i}\left(\sum_{i=1}^{j} w_i q_i\right)$$

Um die Gewichtung zu veranschaulichen, soll ein weiteres Beispiel dienen. Angenommen, eine Lösung erhält vier Bewertungen *($w_1$ = 0,9)*, *($w_2$ = 0,2)*, *($w_3$ = 0,4)* und *($w_4$ = 0,5)* von Teilnehmern, deren Lösungen selbst systeminterne *Qualitätswerte (s.u.) ($q_1$ = 0,8)*, *($q_2$ = 0,1)*, *($q_3$ = 0,3)* und *($q_4$ = 0,4)* haben. Dann ergibt sich für die bewertete Lösung eine *Evaluierungsbewertung e* von:

$$e = \frac{1}{\sum_{i=1}^{4} q_i}\left(\sum_{i=1}^{4} w_i q_i\right) = \frac{1}{1{,}6}\left(0{,}9 \cdot 0{,}8 + 0{,}2 \cdot 0{,}1 + 0{,}4 \cdot 0{,}3 + 0{,}5 \cdot 0{,}4\right) \approx 0{,}66$$

Die erste Bewertung wird hier deutlich stärker gewichtet, als die anderen Bewertungen, da der Bewertende im Gegensatz zu den anderen Bewertenden selbst eine hohe Bewertung inne hat und dessen Meinung somit vom Algorithmus als „kompetenter" eingestuft wird.

Abschließend werden die beiden Komponenten im *Qualitätswert* zusammengeführt. Hierbei werden die Ergebnisse der einzelnen Komponenten abhängig von der Anzahl der abgegebenen Bewertungen *p* zu einer Lösung gewichtet. Die Konstante *c* entspricht der Anzahl der von jedem Teilnehmer zu bewertenden Lösungen. Der *Qualitätswert q* ergibt sich, bezogen auf unser Beispiel, in dem gilt *(c=3)* und *(p=4)*, folglich durch:

$$q = \frac{c}{p+c} b + \frac{p}{p+c} e = \frac{3}{4+3} 0{,}88 + \frac{4}{4+3} 0{,}66 \approx 0{,}75$$

Der Einfluss des *Basiswerts* einer Lösung nimmt folglich mit einer zunehmenden Anzahl an Bewertungen ab. Im CITUC-System wird, anders als im Beispiel, ein Wert von *(c=5)* verwendet.

# 5. Studienbeschreibung

Um das CITUC-System auf seine Praxistauglichkeit hin zu untersuchen, wurde die Vorlesung „Wirtschaftsinformatik II: Modellierung von Informationssystemen“ der Technischen Universität Clausthal im Sommersemester 2008 ausgewählt. Zu den Teilnehmern der Vorlesung gehörten Studenten der Wirtschaftsinformatik (Diplom und Bachelor) und der Betriebswirtschaftslehre (Bachelor) in den unteren Semestern.

Das System wurde am 17.06.2008 nach einer kurzen Einführung in der letzten Vorlesungsveranstaltung, die auch als Videoaufzeichnung und PowerPoint-Präsentation abrufbar war, zur Nutzung freigegeben und stand bis zum Klausurtermin am 29.07.2008 online zur Verfügung. Die Teilnahme am System war allen Studenten auf freiwilliger Basis möglich. Um die Studenten zur Systemnutzung zu motivieren, wurden im Abstand von zwei Wochen E-Mail Erinnerungen verschickt. Bis zum 29.07.2008 hatten sich 98 Teilnehmer im System registriert. An der Klausur nahmen 85 Studenten teil. Insgesamt wurden 50 Aufgaben in das System eingestellt. Davon wurden 22 Aufgaben aus dem vorherigen Übungsbetrieb des Semesters übernommen (diese sollten zur Eingewöhnung in das System dienen) und 27 neue Aufgaben explizit zur Klausurvorbereitung vom Übungsleiter eingestellt. Eine Aufgabe wurde von einem Studenten eingestellt. Einige Tage vor der Klausur wurden die Teilnehmer dazu aufgerufen, einen Abschlussfragebogen auszufüllen, in dem sie das CITUC-System bewerten sollten. Diesem Aufruf kamen 29 der 98 Teilnehmer nach.

# 6. Ergebnisse

## 6.1. Auswirkungen der Modifikationen an der Heuristik

Nachdem die Tauglichkeit der Heuristik bereits ausführlich mittels einer Laborstudie nachgewiesen werden konnte [6], musste nur geprüft werden, ob die dort identifizierten Schwächen im Erreichen von extremen Bewertungen (<0.2 bzw. >0.8) durch die getätigten Modifikation am *Basiswert* beseitigt werden konnten. Eine Analyse der 30 schlechtesten Lösungen (gemäß *Qualitätswert*) ergab einen Mittelwert der einzelnen Lösungsbewertungen von m=0,087 (sd=0,034) bzw. m=0,238 (sd=0,179), falls nur die *Basiswerte* berücksichtigt wurden. Zudem dominierten in diesem Set Spamantworten (ca. 83%), die von Studenten einzig dazu eingestellt wurden, um alternative Lösungen einsehen zu können. Dies lässt darauf schließen, dass der Algorithmus diese Art von Spam erfolgreich herausfiltern kann. Die übrigen Lösungsbearbeitungen in der Liste der „30 schlechtesten“ waren allesamt ebenfalls korrekterweise (gemäß Expertenbewertung) als fehlerhaft klassifiziert. Ein ähnliches Bild zeigte sich bei den Lösungsbearbeitungen, die vom System als die besten 30 klassifiziert wurden. Unter ihnen befand sich eine einzige Spamantwort mit einem hohen *Basiswert*, der offenbar auf sehr gutes Raten seitens des Studenten zurückzuführen ist. Die Antwort erhielt jedoch bis zum Schluss keinerlei weitere Bewertungen, sodass einzig der Basiswert für die Bewertung ausschlaggebend war. Die übrigen 29 Antworten erhielten jeweils 5 Bewertungen und waren korrekt als qualitativ hochwertig klassifiziert. Der Mittelwert der *Qualitätswerte* der einzelnen korrekt klassifizierten Lösungsbewertungen betrug m=0,914 (sd=0,025). Der Mittelwert

der jeweiligen *Basiswerte* betrug m=0,747 (sd=0,139). Damit wies unsere Heuristik insgesamt eine bessere Performanz beim Erreichen von extremen Bewertungen als zuvor auf. Zudem lieferte bereits der Basiswert größtenteils akzeptable Ergebnisse (s. [6] für Vergleichsmaßstab).

## 6.2. Nutzungshäufigkeit

Nachdem die qualitative Verbesserung der Heuristik bestätigt wurde, soll im Folgenden die quantitative Nutzung des Systems untersucht werden. Wie in Abb. 4 bzw. Abb. 5 ersichtlich ist, verlagerten sich die Hauptaktivitäten innerhalb des Systems trotz expliziter Aufrufe zu einer frühen Nutzung in die Endphase, d.h. in die ca. 1,5 Wochen vor dem Klausurtermin, und gipfelten am letzten Tag vor der Klausur in einem globalen Maximum sowohl auf Seiten der Logins pro Tag (s. Abb. 4) als auch mit Blick auf die von den Studenten eingestellten Lösungen pro Tag (s. Abb. 5). Die kleinen Ausschläge bei der Nutzung in den ersten Tagen sowie nach etwa 2 Wochen sind durch die geweckte Neugier der Nutzer sowie durch verschickte E-Mail Erinnerungen zu erklären. Eine freiwillige Nutzung des Systems ist damit keine ausreichende Motivation für einen kontinuierlichen Nutzungsprozess, aber ausreichend für eine kurzfristige Klausurvorbereitung.

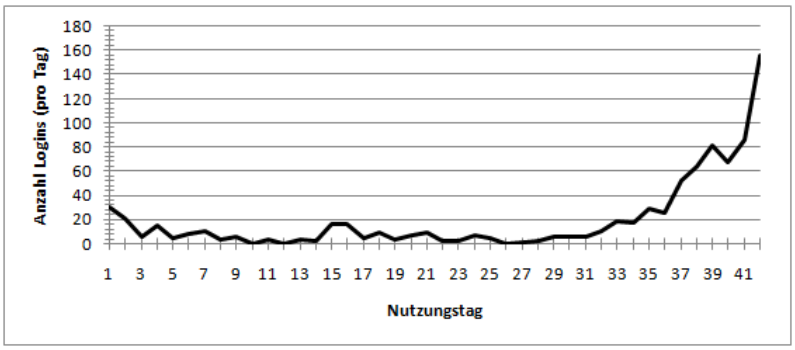


**Abbildung 4: Anzahl der Logins pro Tag (ohne Tutoren / Dozenten)**

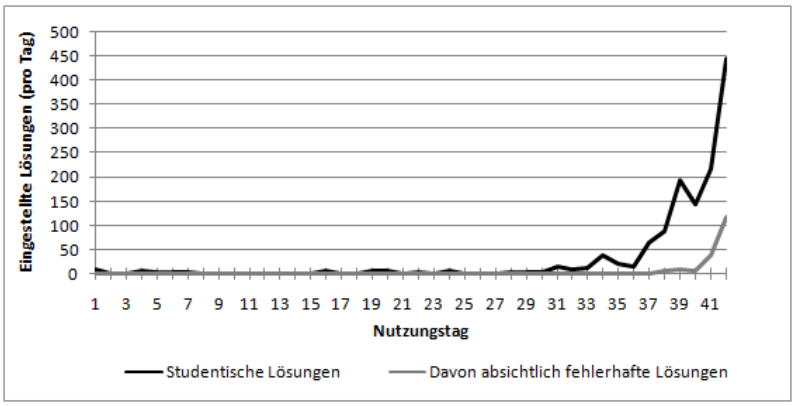


**Abbildung 5: Anzahl der eingestellten Lösungen pro Tag (ohne Vorgaben)**

Wie Abb. 5 verdeutlicht, spielte das System hierbei deutlich seinen Vorteil einer kurzen Bearbeitungszeit im Vergleich zu anderen Ansätzen (PG, SWoRD, s. Abschnitt 2) aus: auch die am letzten Tag eingegangenen Lösungen wurden überwiegend noch bewertet und kommentiert, und erlaubten den Studenten daher Rückmeldungen zu ihren Lösungen bis (fast) zur „letzten Minute“. Abb. 5 zeigt ebenfalls, dass insbesondere in den letzten Tagen der Systemnutzung zunehmend absichtlich fehlerhafte Lösungen eingetragen wurden. Dies ist dadurch erklärbar, dass alternative Lösungen erst nach erfolgter eigener „Aufgabenbearbeitung“ (auch grob fehlerhafter) einsehbar

waren. Auffällige Nutzer wurden zum Teil von den Tutoren durch private Nachrichten zur Ordnung gerufen, was jedoch nur bedingt erfolgreich war. Insgesamt wurden 617 Kommentare (davon 457 vom Dozenten bzw. Tutoren) und 1431 Lösungen (davon 40 Vorgaben) in das System eingestellt.

### 6.3. Einschätzung des Systems durch die Nutzer

Eine Auswertung der Fragebögen ergab, dass die Studenten das System überwiegend als nützlich empfanden. Sie gaben eine Wertung von m=3,89 (sd=0,766, n=26) auf einer Skala von 1 (sehr nutzlos) bis 5 (sehr nützlich) ab. Ein ähnliches Bild ergab sich bei der Frage nach dem Nutzen der Kommentarfunktion, bei der sich Werte von m=3,556 (sd=0,974, n=27) ergaben. Auf die Frage, ob das CITUC-System eine sinnvolle Klausurvorbereitung darstellt, stimmten 18 Teilnehmer für „ja“ und 3 für „nein“, wobei letztere das System nachweislich nicht genutzt haben, d.h. sie haben sich lediglich registriert und die Aufgaben durchgelesen. Zur Frage, wie schwierig die Bedienung des Systems sei, ergab sich eine durchschnittliche Bewertung von m=3,704 (sd=0,993, n=27) auf einer Skala von 1 (sehr schwierig) bis 5 (sehr einfach). Einschränkend hierbei ist festzuhalten, dass die Testgruppe ausschließlich aus Wirtschaftswissenschaftlern und -informatikern bestand. Eine Verallgemeinerung der Ergebnisse auf andere Gruppen ist daher nicht ohne weiteres möglich [12] und sollte in zukünftigen Untersuchungen besondere Beachtung finden.

Ein abschließendes Interview mit dem Übungsleiter ergab, dass dieser den Arbeitsaufwand gegenüber herkömmlichen Übungsformen als gleich einschätzte. Ein von ihm genannter Vorteil war die Möglichkeit, jederzeit Aufgaben in das System einzustellen, um damit „auf in der Veranstaltung erkannte Schwächen nochmals einzugehen“ [Anm.: Es gab auch nach der letzten Vorlesung noch Präsenz-Übungen]. Zudem konnten über CITUC mehr Aufgaben gestellt werden, als in einer üblichen 90-minütigen Übungsveranstaltung zur Klausurvorbereitung möglich gewesen wäre, was unter Berücksichtigung des gleichgebliebenen Arbeitsaufwandes vom Übungsleiter als äußerst positiv angesehen wurde. Wünschenswert sei nach Meinung des Übungsleiters zudem ein vorlesungsbegleitender Einsatz des Systems. Bedenken seinerseits bestanden lediglich hinsichtlich der Frage, ob Studenten auch komplexere Aufgabenlösungen ernsthaft bewerten würden. Dies wurde in der Studie nicht erfasst, da fast alle Aufgabenbearbeitungen kürzerer Natur waren, was zugleich einen möglichen Ansatzpunkt für zukünftige Untersuchungen bietet.

Auf Seite der Studenten wurde deutlich, dass Teilnehmer, trotz ausführlicher Erklärungen vor Beginn der Systemnutzung, vereinzelt den Sinn des CITUC-Systems nicht auf Anhieb verstanden haben und noch stark im traditionellen Modell „Tutor korrigiert Lösungen und stellt Musterlösungen vor“ dachten. So wurde teilweise nach Musterlösungen verlangt, die größtenteils in Form einer sehr hoch bewerteten Systemvorgabe (die Einstufungen der Vorgaben wurden auch nach Evaluierungen nicht geändert) bereits im System vorhanden waren, aber nicht explizit mit dem Wort „Musterlösung“ gekennzeichnet waren. Auch herrschte Misstrauen gegenüber den ermittelten Qualitätswerten, die *teilweise* ignoriert bzw. erst nach einem Kommentar durch einen Tutor, der die Richtigkeit der Systembewertung bestätigte, beachtet wurden. Entgegen der Erwartungen wurde die Möglichkeit, selbst Fragen zu nicht verstandenen Themen im System zu stellen, nur einmal genutzt. Auch hier gilt es, den Grund in zukünftigen Untersuchungen herauszufinden.

## 7. Zusammenfassung und Ausblick

Das in diesem Artikel vorgestellte CITUC-System setzt kollaborative Filteralgorithmen und Peer Reviews im E-Learning ein und wurde zur Klausurvorbereitung an der Technischen Universität

Clausthal erfolgreich eingesetzt. Probleme ergaben sich in der Motivation der Studenten zur kontinuierlichen und sinnvollen Nutzung des Systems bei freiwilliger Teilnahme – vor der Klausur wurde das System jedoch intensiv verwendet und sowohl durch den Übungsleiter als auch durch die Studenten als hilfreich angesehen. Das vorsätzliche Einstellen von sinnlosen Lösungen mit anschließender zufälliger Bewertung der alternativen Lösungen (mit dem Ziel, so schnell wie möglich Alternativlösungen sehen zu können) stellte eine Herausforderung für die Heuristik dar, die jedoch bestanden wurde. In zukünftigen Untersuchungen gilt es zu klären, ob eine verpflichtende vorlesungsbegleitende Nutzung des Systems oder z.B. eine Funktion zum Überspringen von Aufgabenbearbeitungen hier Abhilfe schaffen kann, oder ob das dazu führt, dass gar keine Lösungen mehr eingestellt werden. Zudem wird eine Ausweitung der Systemnutzung über die Klausurvorbereitung hinaus angestrebt, um weitergehende Erkenntnisse zu gewinnen.

## 8. Literatur


[1] CHO, K., SCHUNN, C. D. (2007). Scaffolded Writing and Rewriting in the Discipline: A Web-Based Reciprocal Peer Review System. Computers & Education, Vol. 48 (3).

[2] DANCER, W. T., DANCER, J. (1992). Peer Rating in Higher Education. J. of Education for Business, 67, 306-309.

[3] GEHRINGER, E. F. (2001). Electronic Peer Review and Peer Grading in Computer-Science Courses. In Proceedings of the 32nd SIGCSE Technical Symposium on Computer Science Education, February 2001, Charlotte, North Carolina, United States, pp. 139-143.

[4] GOLDBERG, D., NICHOLS, D., OKI, B. M., TERRY, D. (1992). Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM 35 (12): 61-70.

[5] HINDS, P. J. (1999). The Curse of Expertise: The Effects of Expertise and Debiasing Methods on Predictions of Novice Performance. Journal of Experimental Psychology: Applied, 5, 205-221.

[6] LOLL, F., PINKWART, N. (2009). Using Collaborative Filtering Algorithms as eLearning Tools. In Proceedings of the 42nd Hawai'i International Conference on System Sciences. Hawaii (USA).

[7] LYNCH, C., ASHLEY, K., ALEVEN, V., & PINKWART, N. (2006). Defining Ill-Defined Domains; A Literature Survey. In Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains at the 8th International Conference on Intelligent Tutoring Systems (pp. 1-10). Jhongli (Taiwan), National Central University.

[8] MALTZ, D. und EHRLICH, E. (1995). Pointing the Way: Active Collaborative Filtering. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

[9] MATHEWS, B. (1994). Assessing Individual Contributions: Experience of Peer Evaluation in Major Group Projects. British Journal of Educational Technology, 25, pp. 19-28.

[10] PINKWART, N., ALEVEN, V., ASHLEY, K., & LYNCH, C. (2007). Evaluating Legal Argument Instruction with Graphical Representations Using LARGO. In Proceedings of the 13th International Conference on Artificial Intelligence in Education (p. 101-108). IOS Press.

[11] SUROWIECKI, J. (2004). The Wisdom of the Crowds. Doubleday.

[12] WILDE, T., HESS, T., HILBERS, K. (2008). Akzeptanzforschung bei nicht marktreifen Technologien: typische methodische Probleme und deren Auswirkungen. In Proceedings of Multikonferenz Wirtschaftsinformatik, MKWI 2008, pp. 1031-1042