

Educational Adaptive Hypermedia meets Computer Assisted Assessment^{*}

Enrique Alfonseca, Rosa M. Carro, Manuel Freire, Álvaro Ortigosa, Diana Pérez, and Pilar Rodríguez

Computer Science Department, Universidad Autonoma de Madrid,
Carretera de Colmenar Viejo, km. 14,5,
28043 Madrid, Spain

{Enrique.Alfonseca,Rosa.Carro,Manuel.Freire,Alvaro.Ortigosa,
Diana.Perez,Pilar.Rodriguez}@ii.uam.es

Abstract. In this paper we explore the many possibilities that arise when we combine adaptive web-based courses with computer-assisted assessment. We argue that this integration has several advantages, such as the feasibility of getting a better model of the student's progress, which will be used with adaptation purposes, and the possibility of proposing and evaluating open-ended questions in the way that is judged more suitable for each student.

1 Introduction

Adaptive hypermedia has been widely used for the development of adaptive Web-based courses, in which each student is individually guided during the learning process [1]. Most of these systems obtain feedback from the student from two sources: their behaviour browsing the course (e.g. pages visited, time spent in each or navigational path) and test questions (e.g. true-false, multiple-choice or fill-in-the-blank questions). Some authors have expressed their concern that this limited way of assessment may not be really measuring the depth of the student learning [2]. This fact has been the motivation of the field known as Computer-Assisted Assessment (CAA) of student essays. CAA of student essays is a long-standing problem that has received the attention of the Natural Language Processing research community. There are many possible ways to approach this problem, including a study of the organisation, sentence structure and content of the student essay [3, E-rater], pattern-matching techniques [4, IEMS], or Latent Semantic Analysis [5, IEA]. Valenti et al. [6] describe the state-of-art of CAA systems.

In order to support adaptive distance teaching and learning, we have developed the TANGOW system, which supports the description of adaptive web-based courses and their dynamic generation, so that their components are tailored to each student at runtime [7,8]. We have also developed, independently, a

^{*} This work has been sponsored by CICYT, project number TIC2001-0685-C02-01.

CAA system called Atenea [9] which is based on n-gram co-occurrence metrics [10]. In this paper we describe the ongoing integration of TANGOW and Atenea and the benefits of this integration.

The paper is structured as follows: in Section 2 we describe separately Atenea, TANGOW, and their integration. Section 3 will highlight the conclusions we can already draw from the work performed, and future work.

2 Atenea and TANGOW

Atenea [9,11] is a Computer-Assisted Assessment system for automatically scoring short student answers. It currently uses statistically-based evaluation procedures. The precision of the student's answer is obtained calculating an n-gram co-occurrence metric [10] with a set of references written by the teacher, and its recall is calculated studying the percentage of those references that is covered by the student's answer.

The scores provided by Atenea are values between 0 and 1, where the upper and the lower bounds depend on the particular question. Nonetheless, we can perform a linear regression to find a correspondence between the interval of possible Atenea's scores and a fixed rank of scores, for instance, between 0 and 10. As expected, the quality of the assessment is very influenced by the kind of question and the references written. When scoring definitions, the correlation between Atenea's scores and the marks assigned manually can reach 0.80 [9]. On the other hand, questions that require that the student make a reasoning or compare several topics are more difficult to evaluate with keyword-based procedures.

The system has been tested with ten short questions, which are grouped in three collections: Operating Systems, Advanced Operating Systems and Object-Oriented Programming. Concerning the user models, Atenea is currently configured to show different questions depending on the student's language and experience. Besides, these factors affect the set of reference answers that is chosen.

The feedback that the students get from the system is a numerical score and a copy of their answer where, with colour codes, they can observe which were the coincident n-grams and which words did not appear in any reference. From that output they can easily guess which are the portions of their answers that are correct and have contributed in incrementing their score. Figure 1 shows an example answer page. In the user profile, students may also indicate whether they just want the score and are not interested in receiving this feedback.

The TANGOW system delivers adaptive web-based courses, and has evolved significantly since [7]. Courses delivered by TANGOW are composed of several tasks, that can be accomplished by the students. A task can correspond to either a theoretical explanation, an example, an exercise to be done individually or an activity to be performed collaboratively (problems to be solved, discussions, etc). The set of available tasks is constantly regenerated, tracking changes in the student's profile (static features and dynamic actions). Once a task is chosen, the

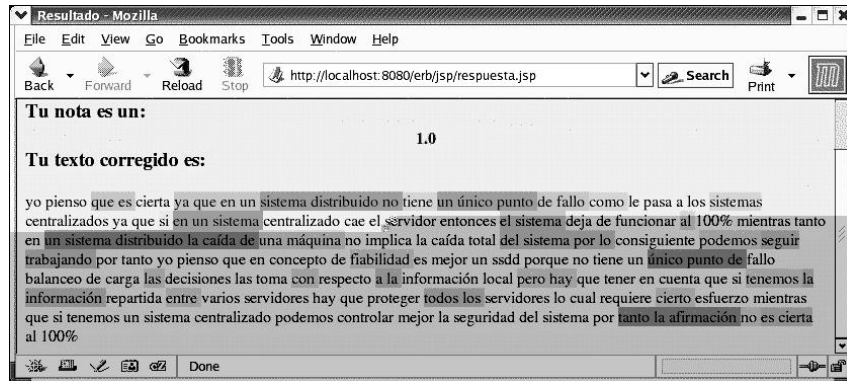


Fig. 1. Feedback that a student gets after answering the question “*Discuss whether distributed systems are more robust than monolithic systems*”, in Spanish. The darker the background, the longer the coincident n-gram.

system generates the corresponding web pages by selecting, among the content fragments and the set of available collaborative tools, those that provide the best possible fit to the current profile.

A rule-based formalism has been developed in order to facilitate the specification of alternative structures for the same course, and to support different teaching strategies, navigational guidance variations and collaboration workspaces for each type of student [8].

The integration of Atenea and TANGOW will support the inclusion of CAA exercises inside adaptive courses. A first consequence will be a richer set of activities, which can contribute to a more engaging learning process. Secondly, the use of the TANGOW formalism allows course authors to specify different teaching strategies by incorporating CAA activities at different points of the course, depending on each student’s evolution. It will be possible for authors to choose the types of users to whom CAA activities will be presented; the places in the course where these exercises will appear; the requirements for a CAA activity to be proposed; and the grading criteria to determine the degree of success of each activity. Each of these adaptations can be made in different ways depending on the user’s model. Finally, the formalism also supports the adaptation of CAA activities themselves: the questions to be asked and the reference answers can be chosen according to each student’s profile.

The results obtained by a student while performing CAA exercises can also be used for adaptation purposes along the rest of the course, without the need of any human intervention to correct the exercises during the course delivery.

3 Conclusions and future work

We are implementing the integration of the adaptive hypermedia educational system TANGOW with Atenea, a program for automatic assessment of student

answers. Atenea attains a very good correlation with respect to teachers' marks, particularly when evaluating definitions and short descriptions [9]. The current implementation allows the students to try out their knowledge, and its complete integration with TANGOW will support the following:

- Atenea will use the description of the user profiles maintained by TANGOW, so it will accept variable profiles.
- The adaptation engine from TANGOW will decide which are the questions that are more adequate for each student, depending on his or her profile, knowledge, and actions, resulting in a fairer evaluation.
- TANGOW will benefit not only from the possibility of automatically evaluating free-text answers, but also from the feedback from those questions, which can be used to guide the students during the rest of the course.
- It will be possible to obtain a dataset of student answers related to their profile and performance in the course, which we shall use in further studies to analyse how the adaptation can improve CAA activities.

The interaction protocol between Atenea and TANGOW has already been designed. Near-future work comprises a complete integration of Atenea and TANGOW along the lines described above, and the evaluation of the integrated system with real students.

References

- [1] Brusilovsky, P.: Adaptive hypermedia. *User Modelling and User-Adaptive Interaction, Ten Year Anniversary Issue* (Alfred Kobsa, ed.) **11** (2001) 87–110
- [2] Whittington, D., Hunt, H.: Approaches to the computerized assessment of free text responses. In: *Proceedings of the Int. CAA Conference*. (1999)
- [3] Burstein, J., Leacock, C., Swartz, R.: Automated evaluation of essay and short answers. In: *Proceedings of the Int. CAA Conference*. (2001)
- [4] Ming, Y., Mikhailov, A., Kuan, T.L.: Intelligent essay marking system. In: C. Cheers (ed.), *Learners Together*, Singapore, NgeeANN Polytechnic (2000)
- [5] Laham, D.: Automated content assessment of text using Latent Semantic Analysis to simulate human cognition. Ph.D. thesis, University of Colorado, Boulder (2000)
- [6] Valenti, S., Neri, F., Cucchiarelli, A.: An overview of current research on automated essay grading. *Journal of I.T. Education* **2** (2003) 319–330
- [7] Carro, R.M., Pulido, E., Rodríguez, P.: Dynamic generation of adaptive internet-based courses. *Journal of Network and Computer Applications* (1999) 249–257
- [8] Carro, R.M., Ortigosa, A., Schlichter, J.H.: A rule-based formalism for describing collaborative adaptive courses. In: Palade, V., Howlett, R., Jai, L. (eds.): *Knowledge-Based Intelligent Information and Engineering Systems. LNAI 2774*. Springer-Verlag (2003) 252–259
- [9] Pérez, D., Alfonseca, E., Rodríguez, P.: Application of the BLEU method for evaluating free-text answers in an e-learning environment. In: *Proceedings of the Language Resources and Evaluation Conference (LREC-2004)*. (2004)
- [10] Papineni, K., Roukos, S., Ward, T., Zhu, W.: Bleu: a method for automatic evaluation of machine translation (2001)
- [11] Pérez, D., Alfonseca, E., Rodríguez, P.: Upper bounds and extension of the BLEU algorithm applied to assessing student essays. In: *IAEA-2004 Conference*. (2004)