

WILLOW: AUTOMATIC AND ADAPTIVE ASSESSMENT OF STUDENTS' FREE-TEXT ANSWERS

Diana Perez-Marin

EPS, UAM
Madrid, Spain
diana.perez
@uam.es

Enrique Alfonseca

EPS, UAM
Madrid, Spain
enrique.alfonseca
@uam.es

Pilar Rodriguez

EPS, UAM
Madrid, Spain
pilar.rodriguez
@uam.es

Ismael Pascual-Nieto

EPS, UAM
Madrid, Spain
ismael.pascual
@uam.es

1 Introduction

Computer Assisted Assessment (CAA) is the field that studies how to effectively use computers to assess students' answers. It is the general opinion in the field that just objective testing (Multiple Choice Question, fill-in-the-blank exercises,...) is not enough to fully evaluate the knowledge of a student. Thus, a subfield has attracted a great deal of attention in the last years. It is called free-text CAA and it is possible thanks to the use of Natural Language Processing (NLP) techniques.

Our goal is not to replace teachers or to address all type of questions, but to provide an alternative form of evaluation that focuses on evaluating open-ended questions to help teachers as a double-checkers or as an initial filtering step and students as additional training.

Willow (<http://orestes.ii.uam.es:8080/ateneaAdaptativa/jsp/loginAtenea.jsp>) is a free-text CAA system that is based on the use of the wraetlic tools (Alfonseca, 2003). It differs from other systems in that it does not need any training or to ask teachers to fill in templates in a certain format. In fact, it only requires that per each question there are several different correct answers in plain text. Besides, it is the first adaptive free-text CAA system as it takes into account the information of the student stored in his or her profile and his or her performance during the assessment session to choose the next more suitable question to the student's knowledge level (Perez-Marin et al., 2006). A non-adaptive version of Willow is also available (Atenea).

2 Willow's authoring tool

After this introduction, the demonstration will centre on the Willow's authoring tool (<http://orestes.ii.uam.es:8080/ateneaAdaptativa/>

[jsp/loginEditor.jsp](#)). In Willow, a course is always associated with a collection of questions belonging to several topics. The teacher has to select the collection to use (or to create a new one) and which question to modify or to add a new one. For each question, the teacher is asked to write its statement and correct answers (for each level of experience and group of age), its topic and the level of difficulty.

The correct answers or references can be manually modified or by means of a genetic algorithm. The genetic algorithm can be used whenever there are previous answers of the students (not the first year of the course) as it automatically gathers correct answers from the students to insert them as references. This is to avoid leaving all the responsibility of writing the correct answers on the teachers. Moreover, given that Willow can be used to process both English and Spanish answers and as teachers are not supposed to know both languages, they only have to write the answers in their mother tongue. Afterwards, if a student's answer is written in a language in which there are no correct answers, it will be automatically translated to the language in which the references are written.

3 Willow

Willow is a highly modular system to make easier the introduction of new modules and its configuration. It also has an associated tool (<http://orestes.ii.uam.es:8080/ateneaAdaptativa/jsp/loginConfigurador.jsp>) to choose the NLP techniques from the wraelite tools that will be applied. The results vary according to the selection done, being the best configuration: stemming and removal of closed-class words with Latent Semantic Analysis (Perez et al., 2005). See Figure 1 for an overview of Willow's architecture.



Figure 1. Willow's architecture

The input of the system is the student's answer in plain text. First of all, the student's answer and the references are processed with the selected tools. For instance, if all the techniques have been selected, the text will be stemmed, the closed-class words will be removed and the sense will be identified. The processed texts enter the comparison module that applies the ERB algorithm (Perez et al., 2005) to them to obtain not only the score to give the student but also the text with a color schema in which a green background indicates that it is good point in the student's answer as it has been matched with some reference. A term identification module is also applied to the references to underline the relevant terms not used by the student. With this information the feedback page (see Figure 2 for an example of Willow feedback's page) is generated to the student but it is only shown when the student passes the question. If the student has failed it, then a compensation dialogue starts in which Willow starts asking him or her up to three questions from a more general to a more specific level of detail to guide the student to the correct answer. Moreover, if a question is not passed the first time it is shown, the next time the processed answer will also be presented to help the student to answer. Next, the student is asked another question in his or her level of difficulty, language, level of experience (e.g., novice or advance) and group of age (child or adult) and the answer will be processed as explained above. It is done until a certain number of questions or amount of time indicated by the student or until he or she passes all the questions.

As the student answers more and more questions, Willow keeps track of how the student is using the relevant terms extracted

from the references. Each term corresponds to a concept. Three different types of concepts have been identified: basic-concepts (BCs, e.g. *process*), topic-concepts that group basic-concepts (TCs, e.g. *concurrency*) and area-of-knowledge-concepts that group topic-concepts (ACs, e.g. *Operating System*). Each concept has associated a confidence-value calculated by Willow that indicates how well the system believes that the student knows that concept according to how he or she uses it in his or her answers. From this information the conceptual model of the student is created and represented as a concept map to the teacher so that he or she can easily see which concepts have been misunderstood and should be reviewed.

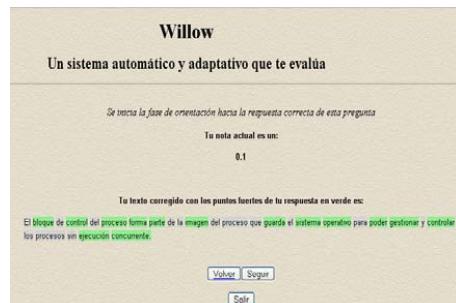


Figure 2. Snapshot of Willow feedback's page.

4 Acknowledgment

This work has been sponsored by Spanish Ministry of Science and Technology, project number TIN2004-0314.

5 References

- Alfonseca, E. 2003. *Wraetlic User Guide version 1.0*. <http://www.ii.uam.es/~ealfon>
- Perez, D., A. Gliozzo, E. Alfonseca, C. Strapparava, B. Magnini, and P. Rodríguez. 2005. About the effects of combining Latent Semantic Analysis with other Natural Language Processing techniques for free-text assessment. *International Journal Signos*, 38(59):325-343.
- Perez-Marin, D., E. Alfonseca, P. Rodríguez. 2006. On the dynamic adaptation of Computer Assisted Assessment of free-text answers. To appear in the *Proceedings of the Adaptive Hypermedia International Conference* published by Springer-Verlag. To be celebrated in Dublin, Ireland.