

# AUTOMATIC GENERATION OF ON-LINE CONCEPTUAL ASSESSMENT COURSES USING TAGHELPER

Ismael Pascual-Nieto<sup>1</sup> and Diana Perez-Marin<sup>2</sup>

<sup>1</sup>Computer Science Department, Universidad Autónoma de Madrid, Spain

<sup>2</sup>Computer and Language Department I, Universidad Rey Juan Carlos, Móstoles, Madrid, Spain

**Keywords:** Computer Assisted Assessment, Text Processing, Authoring of web-based courses, e-Learning.

**Abstract:** TagHelper is a verbal data analysis application. It is based on the use of the Weka toolkit. It is able to classify sentences as one of a set of categories previously introduced into the system. TagHelper has been used to support data analysis in English, German, and Chinese. TagHelper has been recently extended to support Spanish too. The Will Tools are a set of web-based learning tools able to automatically assess students' free-text answers written in Spanish or in English. In this paper, we describe a new procedure to generate a conceptual assessment course in the format required by the Will Tools automatically from web data using TagHelper in Spanish. The procedure has been successfully implemented, and two different courses have already been generated.

## 1 INTRODUCTION

E-learning is only useful when the course is not a mere transcription of the textbook stored in the computer, but the information has been transformed into real knowledge (Simoff & Maher, 1997).

The task of designing a new e-learning course is quite complex. Moreover, the content of the course has to be constantly updated in order to be useful.

In this paper, we are going to describe how the web can also be used to facilitate the automatic generation of courses.

In particular, we will focus on the description of a new procedure to automatically generate courses that can be used in the Will Tools (Perez-Marin et al. 2007), a set of web-based learning available at [www.wisdicor.com/willtools](http://www.wisdicor.com/willtools) to automatically score students' free-text answers.

The procedure has been implemented and two different courses have successfully being created in Spanish.

The paper is organized as follows: Section 2 briefly describes the TagHelper tool; Section 3 briefly describes the Will Tools; Section 4 details the procedure to automatically generate the courses from web data to the Will Tools using TagHelper; and, finally Section 5 ends with the main conclusions and lines of future work.

## 2 TAGHELPER

TagHelper is a verbal data analysis tool (Rosé et al. 2008). It is able to classify a sentence as one of a set of categories previously indicated using the Weka toolkit (Witten & Frank, 2005). That is, it is based on the use of Machine Learning algorithms to induce rules based on patterns found in structured data representations.

In order to achieve this task, TagHelper requires an initial training step. The input for the training step is an Excel file in which each row is a sentence that has been classified by a human rater as one of the categories of the classification. Internally, each row of the input file will be converted into what is known as an instance inside of Weka (i.e. a data point composed of a list of attribute-value pairs). The output of this training step is a model that will be required as one of the inputs of the analysis.

The creation of the model can be configured both in the machine learning algorithm to be used, and in the way in which the structured representation of the text is manipulated.

Finally, it is also important to highlight that the choices of the configuration options and the machine learning algorithm are not entirely independent of one another, and an iterative process can be done to achieve optimum classification results.

### 3 THE WILL TOOLS

The Will Tools (Perez-Marin et al. 2007) are a set of web-based learning tools able to automatically score and provide immediate feedback for short free-text students' answers written either in Spanish or in English, aimed to provide formative assessment.

A course in the Will Tools consists of a set of lessons. In fact, it is recommended to the teachers that they follow the syllabus of their courses.

The first step to create a course in the Will Tools has traditionally been to ask the teachers to introduce the course in the authoring tool. This could be done on-line or uploading the information in a plain-text template as shown in Figure 1.

```

LANGUAGE: [Spanish|English]

COURSE: The name of the course
DESCRIPCION: Optional description of the course

LESSON: The name of the first lesson

Question [maximum_score level_of_difficulty]

- correct answer 1
- correct answer 2

```

Figure 1: Template for a course in the Will Tools.

As can be seen, first of all, it is necessary to indicate whether the content of the course will be in Spanish or in English. Secondly, it is required the name of the course, together with a short optional description. Thirdly, each lesson is described. It can be added as many lessons as needed, all of them following the template.

For each question, it is necessary to indicate its statement, maximum score, level of difficulty (low, medium or high), and a set of correct answers (it can be just one, the answers should be written in natural language and with about a paragraph length).

The correct answers are necessary as the automatic evaluation of the students' short answers work by comparing the answer provided by the student to these correct answers provided by the teachers. The more similar they are, the higher that the system scores the student's answer.

This procedure is standard in free-text Computer Assisted Assessment (Valenti et al. 2003).

Once the course has been introduced into the Will Tools, it can be used by any student from any computer connected to Internet at any time.

Furthermore, students get immediate adaptive feedback after they answer each question.

### 4 PROCEDURE

1. Retrieve data from the web.
2. Use TagHelper (Rosé et al. 2008), a verbal data analysis application to classify the sentences retrieved as definitions or non-definitions. TagHelper can be freely downloaded from [www.cs.cmu.edu/~cprose/THDownload.html](http://www.cs.cmu.edu/~cprose/THDownload.html)
3. Generate the template required by the web-based learning tools to accept the course. In our case, as we have initially used the procedure to generate courses for the Will Tools, we have created the TemGen program that accepts as input the output of TagHelper, and produces the template required by the Will Tools. Nevertheless, provided that a different template was needed, the only change would be to replace TemGen for a different program. A sample pseudo code for TemGen could be as follows:

- 3.1. Create an output.txt file with the header:

```

LANGUAGE: Spanish
COURSE 1st argument of TemGen
DESCRIPTION: 2nd argument
LESSON: Review

```

- 3.2. While there are sentences  $s_1 \dots s_n$  in input.txt, store them in an array of sentences that could be called *arraySentences*.

- 3.3. Group the sentences in *arraySentences* that start by the same noun in groups that could be called *groupSentences*.

- 3.4. For each group discovered in *arraySentences*, write the following sentences in output.txt:

```

Define concept. [1 low]
- first sentence of groupi
- second sentence of groupi
- n sentence of groupi

```

where  $i \in \{1 \dots \text{total number of groups}\}$ , and as can be observed the total number of groups is exactly the same that the total number of concepts, as there is a group of sentences per concept.

4. Upload the generated template in the authoring tool of the Will Tools.
5. The teacher can check the generated course. Initially, each concept is associated to a question. However, the teacher can modify any aspect of the course in case that s/he considers it necessary.



Figure 2: Sample course automatically introduced in the Will Tools from TemGen.



Figure 3: Sample snapshot of a question of the automatically generated course in the Will Tools.

6. Students can log into Willow and start answering the questions, as with any other course of the Will Tools.

Figures 2 and 3 shows snapshots of the sample course generated using this procedure in Spanish.

## 5 CONCLUSIONS AND FUTURE WORK

In the last three decades, it has been produced more information than in the last five millennia. 90% of the information processed by organizations is verbal and, the web is constantly increasing the number of web sites created.



In the 90s, researchers started to consider the use of the web as an educational resource. The benefits are many: to permit access to the course from any computer connected to Internet, at any time and during all the time needed, and to easily update the content of the course without being necessary to send new CDs or books.

However, with all the existing information, the task of creating new web-based courses is more and more challenging. Moreover, taking into account that just typing traditional courses to become e-learning is worthless because it does not provide any added value.

Therefore, some procedures have started to appear to permit the automatic generation of on-line courses from free text such as the ones implemented in PERSEUS (Macias & Castell, 2001) or Welkin (Alfonseca et al. 2004).

However, those approaches do not integrate the possibility of generating open-ended questions with their correct answers to permit their automatic assessment. In this paper, it has been confirmed the possibility of automatically generating conceptual assessment courses. The assessment of concepts is essential in any domain as stated by the Meaningful Learning Theory (Ausubel et al. 1978).

In particular, the procedure has been described using TagHelper and the Will Tools to generate conceptual review courses. However, we believe that it can be extended to other web-based learning tools in which there is a conceptual review section.

As future work, we plan to analyze the generated courses to find out possible improvements to the procedure. Furthermore, we intend to permit the generation of review courses not only for individual concepts, but also to ask for relationships between the concepts as extracted from the web data; and, to extend the procedure to other languages.

In particular, the idea of the procedure can be extended to any other language that can be processed by TagHelper (currently, English, German or Chinese).

Finally, we would also like to carry out an experiment in which a group of students can use the automatically generated courses on a voluntary basis.

## ACKNOWLEDGEMENTS

This work has been sponsored by the projects TIN2007-64718 and CCG08-UAM/TIC-4425.

We would like to thank Carolyn Rosé and Yi-Chia Wang for their explanations about the use of

TagHelper, and to permit the free use of TagHelper, which is a key component of the procedure proposed in this paper.

We would also like to thank you Alvaro Labella and Victor Moreno for their implementation of the TemGen procedure.

## REFERENCES

- Alfonseca, E., Pérez, D., Rodríguez, P., 2004. Welkin: automatic generation of adaptive hypermedia sites with NLP techniques, in *Proceedings of the International Conference in Web Engineering*, LNCS 3140, Springer-Verlag.
- Ausubel, D., Novak, J., Hanesian, H., 1978. *Educational Psychology: a cognitive view*, 2nd. ed., Holt, Reinhart and Winston, New York.
- Macias, J. and Castell, P., 2001. An Authoring Tool for Building Adaptative Learning Guidance Systems on the Web. *Lecture Notes in Computer Science: Active Media Technology-AMT*, Spring-Verlag
- Pérez-Marín, D., Pascual-Nieto, I., Alfonseca, E., Rodríguez, P., 2006. Automatic Identification of Terms for the Generation of Students Concept Maps, in *Proceedings of the International Conference on Multimedia and Information Technologies for the Education* (MICTE), 2007-2011.
- Pérez-Marín, D., Pascual-Nieto, I., Alfonseca, E., Anguiano, E., Rodríguez, P., 2007. A study on the impact of the use of an automatic and adaptive free-text assessment system during a university course, *Blended Learning*, Prentice Hall, Pearson Education.
- Rosé, C., Wang, Y., Cui, Y., Arguello, J., Stegmann, K.; Weinberger, A., Fischer, F., 2008. Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning, *International Journal of Computer-Supported Collaborative Learning* 3(3), 237-271.
- Simoff, S., Maher, M., 1997. Web-mediated courses: The revolution in on-line design education, in *Proceedings of the AusWeb conference*, 143-154.
- Valenti, S., Neri, F., Cucchiarelli, A., 2003. An Overview of Current Research on Automated Essay Grading, *Journal of Information Technology Education* 2.
- Witten, I.H., Frank, E., 2005. *Data mining: Practical machine learning tools and techniques* (2nd ed.), Elsevier: San Francisco.