Título: AUTOMATIC EXERCISE GENERATION BASED ON CORPORA AND NATURAL LANGUAGE PROCESSING TECHNIQUES

Nombre: ALDABE ARREGI, ITZIAR

Universidad: Universidad del País Vasco/Euskal Herriko Unibertsitatea

Departamento: Lenguajes y sistemas informáticos

Fecha de lectura: 20/10/2011

Mención a doctor europeo: concedido

Programa de doctorado: HIZKUNTZAREN AZTERKETA ETA PROZESAMENDUA

Dirección:

> **Director:** MONTSERRAT MARITXALAR ANGLADA

Tribunal:

> **presidente:** RUSLAN MITKOV

> **secretario:** GERMAN RIGAU CLARAMUNT

> **vocal:** HORACIO RODRÍGUEZ HONTORIA

> **vocal:** JENNIFER FOSTER

> **vocal:** GLORIA CORPAS PASTOR

Descriptores:

> INFORMATICA

> ENSEÑANZA CON AYUDA DE ORDENADOR

El fichero de tesis ya ha sido incorporado al sistema

> 2011aldabautom.pdf

Localización: BIBLIOTECA CENTRAL

Resumen: Nowadays, the introduction of information and communication technologies (ICT) into educational areas is a reality. In fact, ICT competence is one part of various curricula. In 2008, even UNESCO published ICT competency standards for teachers. Various official institutions are investing money in introducing technology into classroom. Evidently, the use of ICT is not exclusive to classrooms. For instance, ICTs are widely used in distance learning scenarios. In fact, nowadays, no understanding of distance learning exists which does not involve ICTs.

Thus, ICTs are widely used in different scenarios as media and methodologies. In this dissertation, we present ICT as an approach to help in the learning process of certain subjects. It is undeniable that the effort put into the creation of didactic resources and contents leads to great results as regards their pedagogic appropriateness. In contrast, a large amount of this data is static, and after a certain period of time, could become outdated.

The analysis of various available natural language processing (NLP) tools and corpora has demonstrated that it is possible to implement a system that helps experts and teachers in the creation of didactic material. Thus, we have designed and implemented a system called ArikIturri that, based on NLP and corpora, is able to produce items of a certain standard. ArikIturri is a multilingual system, and different question types have been tested in several scenarios. We have proven the viability of the system to work in the Basque language learning, English language learning and science domains. The experiments have corroborated the feasibility to produce several types of question: error correction, fill-in-the-blank, word formation, multiple-choice and short answer questions. The representation of the items as well as the information relating to their generation process is carried out by means of a question model. This structured representation allows the importation and exportation of the items into independent applications.

We have conducted various experiments in which distinct linguistic information has been utilised. In our experiments, the input for the system is always a corpus, from which sentences are selected to be part of the items based on diverse criteria. In addition, their grammatical and semantic information enabled us to carry out experiments: (i) to prove the viability of the system designed to implement a complete automatic process to generate items; (ii) to apply different methods in the generation of distractros; and (iii) to modify some components of the source sentences when creating the stems. The results of these experiments were obtained from experts' opinions and students' answers. In this way, a qualitative analysis based on experts knowledge gave us a way of measuring the correctness of the automatically generated questions. In addition, the quantitative analysis based on students' responses ensured the quality of the items.