

Generación de preguntas y respuestas que faciliten la comprensión de los contenidos en la web

Marta Gatius Vila

Departament de Ciències de la Computació
Universitat Politècnica de Catalunya
gatius@cs.upc.edu

Resumen

Este artículo estudia la utilidad de presentar los contenidos de las webs académicas en forma de preguntas y respuestas para facilitar su comprensión. Con esta finalidad, se ha utilizado el prototipo de una herramienta que genera semi-automáticamente preguntas y respuestas. Esta herramienta, desarrollada en trabajos de investigación anteriores, se basa en una apropiada representación del conocimiento conceptual y lingüístico implicado.

En una primera fase del estudio, la herramienta se adaptó para generar preguntas y respuestas sobre los exámenes, controles y trabajos de cursos universitarios. Estas preguntas y respuestas se generaron para tres cursos diferentes, en castellano, catalán e inglés, y fueron incorporadas en la plataforma web utilizada en los cursos.

La utilidad de las preguntas y respuestas generadas se evaluó mediante un cuestionario, que se incluyó en la misma plataforma del curso y que fue completado por los estudiantes de forma anónima. Los resultados indicaron un alto grado de satisfacción por parte de los estudiantes, por lo que se extendió la utilización de la herramienta para generar cuestiones que faciliten el aprendizaje de los conceptos básicos del curso. En concreto, se han generado preguntas y respuestas relacionadas con los primeros temas de un curso de introducción a la programación. Las cuestiones generadas se han incorporado en la web del curso, en forma de cuestionarios, que facilitan al estudiante la comprensión de los conceptos fundamentales.

Abstract

This paper studies the use of question and answers to facilitate the understanding of the contents of academic webs. For this purpose, a research tool to (semi-automatically) generate questions and answers, developed in a previous work, has been adapted to this domain. This tool is based on an appropriate representation of the conceptual and linguistic knowledge involved.

In a first step of this study, the tool was adapted to generate questions and answers about the exams and assignments of university courses. The resulting questions (and their corresponding answers) were generated for three different courses in Spanish, Catalan and English. Then, they were incorporated in the courses web sites.

In order to evaluate the usability of the generated questions and answers a questionnaire about them was incorporated into the course platform and the students were asked to complete it anonymously. The results indicated the students satisfaction was high. Thus, further extensions of the work are being studied to use questions and answers to facilitate the learning of the main concepts of the course. In particular, questions and answers related to the basic concepts of an introductory course on programming were generated. They were included in the web course, as questionnaires that students can use.

Palabras clave

Preguntas y respuestas, facilitar la comprensión del contenido de las web académicas, multilingüismo.

1. Motivación

En los últimos años internet se ha venido utilizando como un recurso básico en la educación. No obstante, la mayoría de los sitios web utilizados no han sido diseñados como un recurso educacional y no incorporan contenidos útiles para aprender que, sin embargo, se pueden encontrar en la mayoría de libros de texto como ejercicios, resúmenes o preguntas.

En este contexto, el desarrollo de recursos que faciliten la generación de las preguntas más relevantes relativas a un contenido web podría ser de gran utilidad. En realidad, la generación automática de preguntas y respuestas se ha estudiado en diferentes áreas, como la educación, la divulgación, la medicina, etc. La mayoría de investigaciones en esta línea se han centrado en las preguntas y respuestas para interfaces que interaccionan con el usuario en

aplicaciones diversas: tutores inteligentes (descritos en [2]), guía de museos (como la interfaz presentada en [3]), atención médica,.. No obstante, existen también investigaciones (como los trabajos presentados en [2, 4]), cuyo objetivo consiste en la generación de preguntas y respuestas que puedan utilizarse como un recurso educativo para profesores y estudiantes (por ejemplo, para incorporarlas en exámenes,...).

Aunque se han realizado trabajos de investigación relevantes en las áreas citadas, en la mayoría de casos, su aplicación ha sido limitada debido a que la utilización de estos sistemas por no expertos, así como su evaluación y su mantenimiento, no es fácil.

En este contexto, el trabajo aquí descrito tiene como objetivo principal el estudio de la aplicación, en estudios universitarios, de un prototipo para generar preguntas y respuestas en varias lenguas, desarrollado en trabajos previos de investigación en el área del procesamiento del lenguaje ([1]). Un objetivo adicional consiste en estudiar cómo la presentación de contenidos en forma de preguntas y respuestas (y/o cuestionarios) relevantes facilitan mejoras en la docencia.

En una primera fase, el trabajo se ha centrado en las cuestiones relativas a los diferentes elementos que se consideran en la evaluación de los cursos (exámenes, controles, trabajos,...), sin duda uno de los temas que despiertan más el interés del estudiante. Esta información acostumbra a estar detallada de forma clara en la mayoría de páginas web que describen cursos universitarios, sin embargo, los estudiantes frecuentemente piden aclaraciones sobre estos datos a los profesores.

Los resultados positivos en la evaluación de la satisfacción de los estudiantes con las primeras cuestiones generadas, han propiciado el estudio de la adaptación del prototipo para generar cuestiones que faciliten el aprendizaje de los conceptos básicos de un curso. En concreto, se han generado preguntas y respuestas relacionadas con los contenidos de un curso de introducción a la programación, y se han incorporado en la plataforma web del curso, en forma de cuestionarios. Los cuestionarios además de facilitar la comprensión de los nuevos conceptos, también permiten al profesor realizar un seguimiento del aprendizaje de los estudiantes, mejorar su participación en clase,...

2. La experiencia docente

Esta sección detalla la adaptación del prototipo para generar las preguntas y respuestas referentes a la descripción del curso, así como su evaluación. La ampliación del estudio para generar cuestiones relativas al temario, se describe de forma más general, aunque el procedimiento para obtener las

preguntas y respuestas más relevantes en un dominio es siempre el mismo. En un primer paso deben representarse los conceptos principales, como un conjunto de atributos y, a continuación, estos atributos se relacionan con las correspondientes estructuras lingüísticas.

El prototipo genera preguntas que se ciñen a datos objetivos, es decir, más fáciles de responder que otros tipos de preguntas, como las que requieren inferencias de algún tipo.

2.1. La representación del conocimiento

La herramienta utilizada se basa en una representación declarativa y separada del conocimiento conceptual y lingüístico implicado. Para su aplicación a un escenario concreto, primero

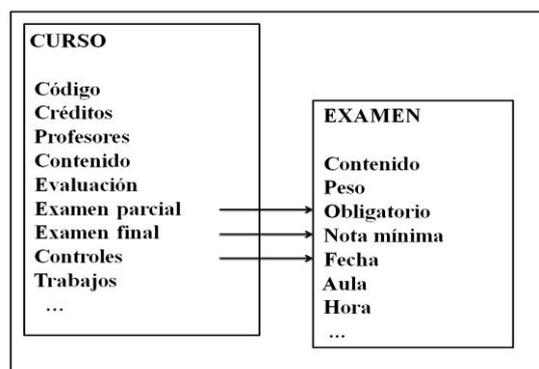


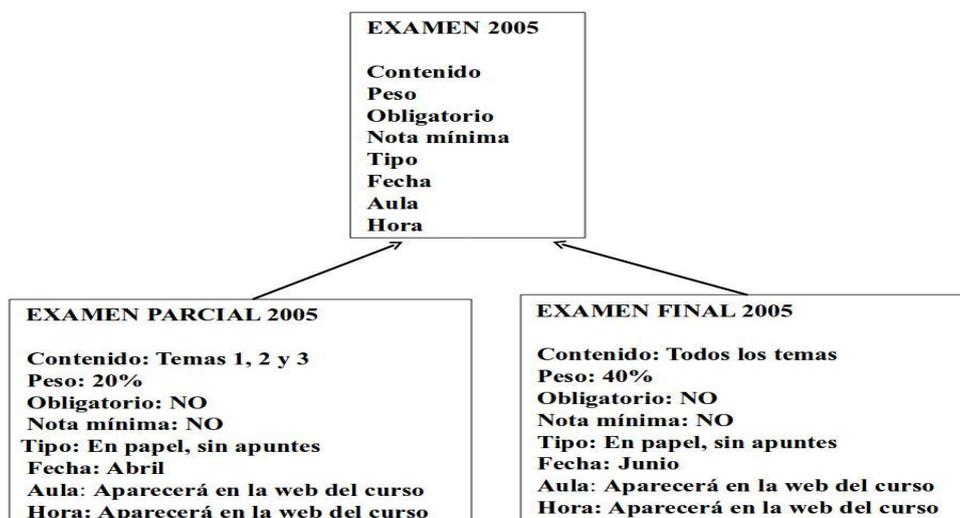
Figura 1: Representación de los conceptos

deben definirse los conceptos principales involucrados, como un conjunto de atributos.

Con la finalidad de adaptar la herramienta al dominio de las descripciones de cursos universitarios, se analizaron diversas web universitarias. Aunque estas descripciones se presentan en formatos diferentes, la mayoría incluyen los datos básicos relativos a contenidos, profesores y evaluaciones.

La Figura 1 muestra la representación (parcial) de los conceptos centrales en el escenario considerado, *Curso* y *Examen*, descritos como un conjunto de atributos. Para cada uno de los cursos considerados se trabajará con instancias o casos concretos de estos dos conceptos generales. Es decir, la definición de cada curso consistirá en el mismo conjunto de atributos, y solamente cambiará su valor.

Como puede observarse en la Figura 1, el concepto *Curso* está descrito, entre otros, por los siguientes atributos: *Código*, *Créditos*, *Profesores*, *Contenido*, *Evaluación*, *Examen parcial*, *Examen final*, *Control* y *Trabajos*. El atributo *Evaluación* detallará la fórmula utilizada para la evaluación del curso. Además, cada examen (parcial, final y controles) es representado por un atributo diferente,

Figura 2: Representación de los exámenes del *Curso 2005*.

cuyo valor será una instancia del concepto general *Examen*, y por ello se definirá con los mismos atributos.

La Figura 2 muestra la representación de los exámenes parcial y final de un curso en particular, identificado con el código *2005*. Como se puede observar, los dos exámenes se definen con el mismo conjunto de atributos, porque se trata de casos concretos (instancias) del mismo concepto.

Descrita la información conceptual del escenario, ésta debe de relacionarse con la información lingüística. Para ello, los conceptos y atributos deben de asociarse con las entradas léxicas correspondientes y, además, cada uno de los atributos con las clases en una taxonomía sintáctico-semántica.

La taxonomía sintáctico-semántica, definida en trabajos de investigación previos ([1]), permite establecer relaciones generales entre el conocimiento conceptual y lingüístico, en diversas lenguas. Las clases en la taxonomía representan atributos para los que las preguntas y respuestas se expresan de forma similar (utilizando las mismas estructuras lingüísticas). Por ejemplo, el atributo *Créditos* pertenece a la clase que expresa cantidades y una pregunta sobre su valor tendrá la forma *¿Cuántos <nombre de atributo> tiene <nombre de concepto>?* (*¿Cuántos créditos tiene el curso 2005?*).

Una vez se han definido los conceptos básicos como un conjunto de atributos, y se han asociado a la taxonomía descrita, se obtienen los valores de los atributos de la página web describiendo el curso, de forma manual o automática. A partir de la representación resultante, se generan las preguntas y respuestas automáticamente, si bien estas deben ser

supervisadas. Este procedimiento se adaptó a tres cursos diferentes.

2.2. Preguntas y respuestas generadas

1. Cuál es el contenido el examen parcial?
Los temas 1,2 y 3.
2. Cuanto vale el examen parcial en la nota final?
El examen parcial vale el 20% de la nota final.
3. Es obligatorio el examen parcial?
No
4. Hay nota mínima del examen parcial?
No
5. Cómo se realiza?
En papel, sin apuntes.
6. Cuándo es el examen parcial?
En abril.
7. A qué hora es el examen parcial?
Se publicará en la web del curso.
8. Dónde es el examen parcial?
Se publicará en la web del curso..
9. Cuál es el contenido el examen final?
Todos los temas.
10. Cuanto vale el examen final en la nota final?
El examen final vale el 20% de la nota final.
11. Cómo se realiza?
En papel, sin apuntes.
12. Cuándo es el examen final?
En junio
- 13- A qué hora es el examen final?
Se publicará en la web del curso.
14. Dónde es el examen final?
Se publicará en la web del curso.
15. Es obligatorio el examen final?
No
16. Hay nota mínima del examen final?
No

Figura 3: Preguntas referentes a exámenes

La Figura 3 muestra preguntas y respuestas generadas en castellano a partir de la representación del examen parcial y final, en la Figura 2. Estas preguntas fueron también generadas en catalán e inglés y se incorporaron en la página web del curso.

Para evaluar la utilidad de las preguntas y respuestas generadas se incorporó un cuestionario en la plataforma educativa utilizada en los cursos de la universidad, *atenea* (basada en *moodle*). En concreto, se utilizó la herramienta incluida en la plataforma para la creación y evaluación de cuestionarios, con la que tanto los docentes como los estudiantes estaban familiarizados. Se pidió a los estudiantes que lo completaran de forma anónima.

La Figura 4 muestra las preguntas incorporadas al cuestionario, a las que los estudiantes debían responder con un valor numérico entre 0 y 10, donde 0 representa totalmente en desacuerdo y 10 totalmente de acuerdo. Las respuestas corresponden a los estudiantes de dos cursos diferentes de introducción a la informática (de dos grados diferentes, de ingenierías no informáticas).

Preguntas	Grado Ingeniería Aeronáutica (sobre 10)	Grado Ingeniería Industrial (sobre 10)
¿Las preguntas y respuestas sobre el curso te han ayudado?	8.10	8.43
¿Crees que es útil su incorporación en la página web del curso?	8.14	8.78
¿Crees que las consultarás otra vez?	9.14	9.78

Figura 4: Evaluación de las preguntas y respuestas

Como puede observarse, los resultados indicaron un alto grado de satisfacción por parte de los estudiantes, lo que incentivó el estudio de utilizar cuestiones que faciliten el aprendizaje de los contenidos básicos del curso de introducción a la programación.

Con la finalidad de simplificar la representación de los conceptos fundamentales se definió un *meta-concepto* con los atributos generales que describen su definición, los elementos que lo constituyen y la definición de cada uno de ellos. Ejemplos de conceptos representados son: *Etapas en el desarrollo de un programa, tipos de datos, operaciones elementales, tipos de operadores, estructuras de control,...*

Las preguntas y respuestas generadas a partir de estos conceptos se han incorporado en la plataforma web del curso, en forma de cuestionarios. En concreto, se ha utilizado la herramienta de *atenea* para presentar las cuestiones en diversos formatos: *verdadero o falso, escoger una o varias respuestas* (de una lista), *emparejar preguntas y respuestas,...* Los cuestionarios han resultado ser un recurso útil para mejorar la participación en clase.

3. Conclusiones y trabajo futuro

En este artículo se estudia la utilidad en los cursos universitarios de un prototipo que facilita la generación de preguntas y respuestas relevantes. En concreto, se ha utilizado para facilitar al estudiante la comprensión de las descripciones en la web de diferentes cursos académicos. Se evaluó la utilidad de las preguntas generadas mediante un cuestionario respondido de forma anónima por los estudiantes de los respectivos cursos, obteniendo resultados muy satisfactorios.

Actualmente, se está estudiando la utilidad de las cuestiones y respuestas generadas para facilitar el aprendizaje de los contenidos de un curso de introducción a la programación. La plataforma docente utilizada ha facilitado la incorporación de las preguntas en cuestionarios en los que se corrige y evalúan las respuestas automáticamente. Estas pruebas además de ayudar a entender los conceptos básicos, facilitan al profesor el seguimiento del aprendizaje de sus estudiantes y motivar su participación en clase,... La evaluación informal de los estudiantes respecto estos cuestionarios es positiva. En esta línea, el trabajo se ampliará para mejorar el proceso de generar estas nuevas preguntas, así como para estudiar y evaluar cómo obtener nuevos tipos de cuestiones.

El trabajo futuro incluirá también el estudio de la posible generación de preguntas personalizadas a estudiantes con diferentes niveles y/o intereses. Debido a que la herramienta utilizada se basa en una representación de los conceptos más relevantes del dominio como un conjunto de atributos, éstos pueden adaptarse fácilmente a los conocimientos (y/o intereses) de diferentes estudiantes.

Referencias

- [1] Marta Gatiús. Personalized Questions, Answers and Grammars: Aiding the Search for Relevant Web Information. *International Conference on Natural Language Generation*, 2017.
- [2] Michael Heilman. 2011. Automatic Factual Question Generation from Text. Ph.D. thesis, Carnegie Mellon University.
- [3] Oliviero Stock, Massimo Zancanaro, Paolo Busetta, Charles Callaway, Antonio Krüger, Michael Kruppa, Tsvi Kuflik, Elena Not, and Cesare Rocchi. 2007. Adaptive, intelligent presentation of information for the museum visitor in PEACH. *User Modeling and User-Adapted Interaction* 17, 3 (July 2007), 257-30.
- [4] Brendan Wyse and Paul Piwek. Generating Questions from OpenLearn study units. *Workshop on Question Generation*, AIED, 2009.