

ADA-Madrid



Relada

(Revista Electrónica de ADA)

Vol. 5 (3) 2011

ISSN: 1988-5822



Trivial: Sistema de Diálogo Hablado para Enseñanza de Contenidos Didácticos Universitarios

**Ramón López-Cózar. Gonzalo Espejo.
Nieves Ábalos. Zoraida Callejas**

Dpto. Lenguajes y Sistemas Informáticos. ETSI Informática y de Telecomunicación.
CITIC-UGR. Universidad de Granada.
rlopezc@ugr.es gonzalp@correo.ugr.es
nyeves@gmail.com zoraida.@ugr.es

David Griol

Universidad Carlos III de Madrid. Departamento de Informática. Escuela Politécnica Superior.
dgriol@inf.uc3m.es

José F. Quesada

Dpto. de Inteligencia Artificial y Ciencias de la Computación. Universidad de Sevilla.
Jose.Quesada@infinity.es

Resumen: Este artículo describe la utilización de las nuevas tecnologías de la información y las comunicaciones para favorecer el aprendizaje autónomo y la autoevaluación de conocimientos adquiridos por alumnos universitarios, siguiendo la filosofía del nuevo Espacio Europeo de Educación Superior. En concreto, se propone el uso de un sistema de diálogo hablado, llamado **Trivial**, cuya finalidad será interactuar vía telefónica con alumnos universitarios para facilitarles el aprendizaje de contenidos didácticos. El artículo presenta la arquitectura del sistema, muestra un ejemplo del diálogo hablado que pretendemos permita el sistema, y describe las tareas que tenemos previsto realizar para implementar el sistema y evaluar su funcionamiento.

Palabras clave: Sistemas de diálogo hablado. Reconocimiento. Comprensión y síntesis de habla. Gestión del diálogo. Voicexml.

Abstract: This paper proposes to use the new information and communication technologies to stimulate self learning and evaluation of the knowledge acquired by university students, in accordance with the guidelines of the new European System for Higher Education. Specifically, the paper proposes to use a spoken dialogue system, called TRIVIAL, to provide university students with speech-based interaction on the phone with the goal of helping them to learn didactic materials. The paper focuses on the architecture of the system, shows an example of the spoken dialogue that we want the system to handle, and

describes the tasks we plan to carry out in order to implement the system and evaluate its performance.

Keywords: Spoken dialogue systems. Automatic speech recognition and synthesis. Spoken language understanding. Dialogue management. Voicexml.

INTRODUCCIÓN

La enseñanza universitaria se ha basado tradicionalmente en el papel desempeñado por el profesor, quien imparte lecciones magistrales de teoría y prácticas. No obstante, el **Espacio Europeo de Educación Superior (EEES)** implica un cambio en la concepción pedagógica, según el cual, se debe pasar de una docencia centrada en la enseñanza (contenidos) a una docencia centrada en el aprendizaje de los alumnos (competencias). Según este nuevo enfoque, la función del docente universitario ya no será tanto impartir la enseñanza como crear las condiciones para el aprendizaje. Así, la enseñanza se constituye en un medio, no un fin en sí misma (Gairín et al. 2004; Troyano et al., 2007). Las competencias representan el conjunto de atributos con respecto al conocimiento y su aplicación, a las actitudes y las responsabilidades de los alumnos, que describen los resultados del aprendizaje de un determinado programa. Por consiguiente, las competencias indican en qué medida los alumnos serán capaces de desenvolverse al final del proceso educativo.

Desde hace algunos años, las **Tecnologías de la Información y las Comunicaciones (TICs)** han comenzado a aplicarse al ámbito de la enseñanza. El objetivo es desarrollar metodologías y herramientas automáticas para afianzar el proceso de aprendizaje de los alumnos, haciéndoles más partícipes en dicho proceso. Un ejemplo es la plataforma Moodle², mediante la cual, los alumnos pueden repasar los contenidos explicados en clase, resolver problemas y ejercicios prácticos, y trabajar colaborativamente. No obstante, y a pesar de sus ventajas, el uso de estos sistemas presenta algunos inconvenientes. Por una parte, su uso exige que los alumnos dispongan de acceso a Internet y de un ordenador para acceder al sitio web donde se encuentran almacenados los materiales didácticos. Por otra parte, este tipo de interacción puede no ser viable para alumnos con problemas de movilidad o visión que puedan limitarles a la hora de usar el teclado o el ratón.

En la actualidad, es cada vez más frecuente que utilicemos nuestra voz para interactuar con sistemas informáticos que ofrecen servicios telefónicos automatizados. Estos sistemas, conocidos coloquialmente como "centralitas automáticas", se conocen técnicamente con el nombre de **sistemas de diálogo hablado** (spoken dialogue systems). Una característica fundamental de estos sistemas es que se diseñan para interactuar con los usuarios mediante un diálogo hablado que sea "inteligente", y que en la medida de lo posible, se asemeje a un diálogo persona-persona (López-Cózar y Araki, 2005). Diversas empresas emplean en la actualidad este tipo de sistemas para proporcionar información a sus clientes de forma automática, por ejemplo, acerca de viajes, cuentas bancarias, conexión a Internet, etc. La ventaja para

² <http://moodle.org/>

las empresas es importante, pues estos sistemas permiten abaratar costes de personal y ofrecer un servicio automático las 24 horas del día, todos los días del año. En particular, estos sistemas permiten **acceder a información en Internet** de una forma cómoda y natural para los usuarios, sin necesidad de usar un teclado o un ratón. Por ejemplo, el sistema Jupiter (Zue et al., 2000) proporciona información meteorológica de un gran número de ciudades de todo el mundo, localizando dicha información en páginas web.

Creemos que esta tecnología puede aplicarse también al ámbito docente universitario, como una nueva herramienta que tienda a mejorar el aprendizaje de los alumnos (García y Lavié, 2000; Ai et al., 2006; Griol et al., 2010). Por este motivo, nos proponemos implementar el sistema de diálogo hablado TRIVIAL, que se describe en la siguiente sección.

SISTEMA TRIVIAL

En este artículo se propone el desarrollo de un sistema de diálogo hablado, al que hemos denominado **TRIVIAL**, cuyo objetivo es favorecer el autoaprendizaje de contenidos didácticos por parte de los alumnos de la Universidad de Granada. La interacción entre los alumnos y el sistema se realizará telefónicamente mediante diálogos hablados. Por tanto, el sistema se diferenciará notablemente de otras herramientas de apoyo a la docencia basadas en TICs. Creemos que el sistema fomentará el desarrollo del espacio innovador promovido por el EEES, pues dará una visión más participativa y atractiva de las asignaturas, proporcionando un entorno innovador que esperamos motive a los alumnos.

Arquitectura del sistema

El sistema se está implementando usando las cinco tecnologías básicas empleadas usualmente para crear este tipo de sistemas, que son las siguientes: Reconocimiento Automático de Habla (RAH), Procesamiento de Lenguaje Natural (PLN), Gestión del Diálogo (GD), Generación de Lenguaje Natural (GLN) y Conversión Texto-Habla (TTS, Text-To-Speech). Para hacer uso de estas tecnologías, implementaremos los módulos del sistema que se muestran en la

Figura 1. Además, el sistema utilizará bases de datos remotas en las que se almacenarán las preguntas que el sistema planteará al alumno junto con sus correspondientes respuestas, con la finalidad de favorecer el aprendizaje de contenidos didácticos.

El flujo de información en el sistema será como se indica a continuación. El módulo de RAH recibirá el habla (voz) del alumno transmitida a través de la línea telefónica, y generará una secuencia de palabras reconocidas (Rabiner y Huang, 1993). Para ello, utilizará modelos acústicos entrenados para el idioma castellano, así como modelos de lenguaje creados para reconocer las frases que se espera sean pronunciadas por el alumno, en su mayoría, respuestas a las preguntas que el sistema le plantee. Este módulo utilizará también un diccionario, que contendrá cada una de las palabras susceptibles de ser reconocidas.

La secuencia de palabras proporcionada por el módulo de RAH constituirá la entrada del módulo de PLN, el cual utilizará reglas gramaticales para generar una representación abstracta que capture el significado de la secuencia de palabras reconocidas (Allen, 1995).

El módulo de GD utilizará la representación semántica creada por el módulo de PLN, la historia del diálogo almacenada en el módulo de Memoria, y una serie de estrategias de interacción, para decidir la siguiente acción que deberá realizar el sistema (Griol et al., 2008). Una posible acción consistirá en consultar las bases de datos a través de Internet para buscar las preguntas que plantear al alumno. Otra posible acción consistirá en solicitar al alumno que confirme determinadas palabras por no estar seguro de que hayan sido reconocidas correctamente. Por ejemplo, el sistema podrá preguntarle: "¿Has dicho disco duro?".

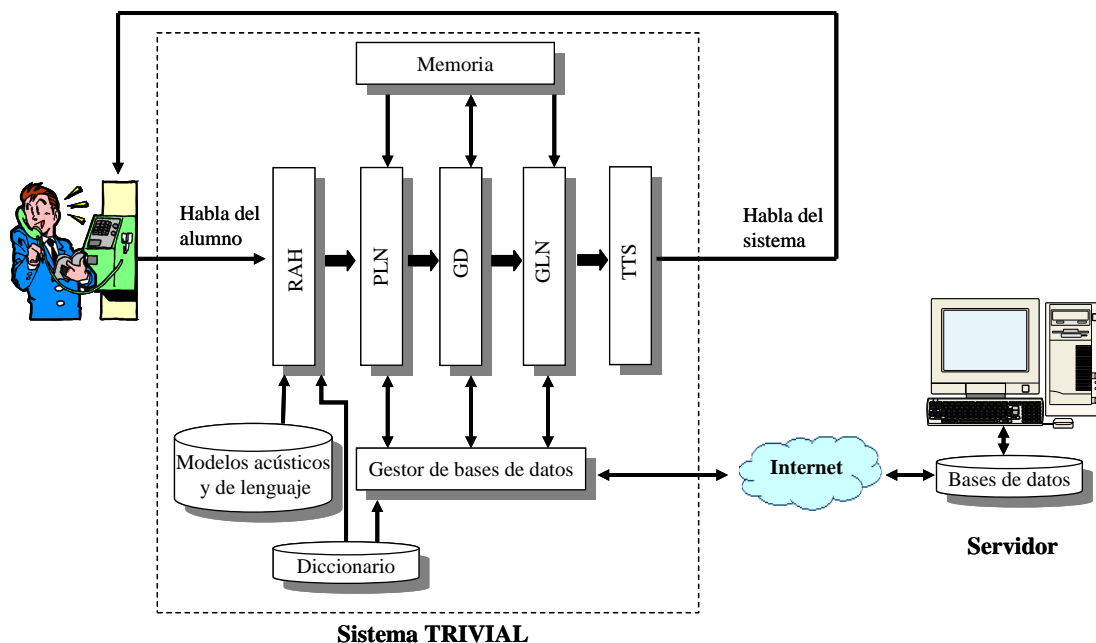


Figura 1. Interacción con el sistema Trivial.

La decisión sobre la siguiente acción del sistema, realizada por el módulo de GD, constituirá la entrada del módulo de GLN, el cual utilizará un conjunto de patrones para generar respuestas en modo de texto, que deberán ser correctas desde punto de vista gramatical y semántico. Finalmente, dichas respuestas se transformarán a voz mediante el módulo de TTS y se enviarán al alumno a través de la línea telefónica.

Ejemplo de diálogo

La figura 2 muestra un ejemplo de diálogo hablado sistema-alumno, según el diseño actual de la interacción. En esta figura, al inicio de cada turno en el diálogo, "S" representa que el turno corresponde al sistema, mientras que "A" indica que el turno corresponde al alumno. Como puede observarse, el alumno tiene un comportamiento espontáneo en la interacción, pues en ocasiones no sabe muy bien qué responder, por ejemplo, en los turnos A1 y

A5. En respuesta a estas situaciones, el sistema le proporciona ayuda (turno S2) o le repite la pregunta que acaba de formular (turno S6).

También se puede observar que en una ocasión, el sistema no está seguro de haber reconocido correctamente las palabras pronunciadas por el alumno, en cuyo caso, solicita una confirmación (turno S7) que el alumno proporciona en el turno A7. Asimismo, puede observarse que el sistema informa acerca de la corrección o incorrección de la repuesta que el alumno ha proporcionado, y en su caso, le informa de la respuesta correcta para favorecer el aprendizaje.

S1: Buenos días. Bienvenido al sistema TRIVIAL. Podrás estudiar contenidos de asignaturas la Universidad de Granada. Dime en primer lugar, a qué Titulación pertenece la asignatura que quieres estudiar.
A1: Pues no me acuerdo ahora de cuál es el nombre exacto de la Titulación.
S2: Las Titulaciones disponibles son: Ingeniería Técnica e Informática de Sistemas, Ingeniería Técnica en Informática de Gestión, Ingeniería Informática, y Licenciado en Traducción e Interpretación.
A2: Licenciado en Traducción e Interpretación.
S3: De acuerdo, Licenciado en Traducción e Interpretación. Por favor, dime qué asignatura quieres estudiar.
A3: Informática Aplicada a la Traducción.
S4: De acuerdo. ¿De qué tema quieres estudiar contenidos?
A4: Del tema uno.
S5: Muy bien. Voy a empezar a hacerte preguntas del tema número 1 para que me digas cual es la respuesta correcta. Atención, comenzamos. Pregunta número 1. De los siguientes elementos del ordenador, indica cual no forma parte de la placa base del ordenador: a) la tarjeta de vídeo, b) la memoria RAM, c) el reloj, d) el disco duro.
A5: ¿Me puedes repetir la pregunta, por favor?
S6: Repito, pregunta número 1. De los siguientes elementos del ordenador, di cual no forma parte de la placa base del ordenador: a) la tarjeta de vídeo, b) la memoria RAM, c) el reloj, d) el disco duro.
A6: Creo que es el disco duro.
S7: ¿Has dicho disco duro?
A7: Sí.
S8: Muy bien, la respuesta es correcta. Continuamos. Pregunta número 2. ¿Cuántos bits tiene un Byte? a) Cuatro. b) Seis. c) Diez. d) Ninguna de las anteriores opciones es correcta.
A8: Creo que seis.
S9: Lo siento, la respuesta correcta es: Ninguna de las anteriores opciones es correcta, pues un Byte tiene 8 bits. Continuamos. Pregunta número 3...

Fig. 2. Diálogo de ejemplo sistema-alumno"

Tareas para el desarrollo del sistema

Para implementar el sistema estamos utilizando el Call Center que tenemos disponible en nuestro laboratorio gracias a un proyecto de investigación previo. El Call Center tiene instalados los motores necesarios para RAH y TTS, así como un intérprete para ejecutar scripts escritos en lenguaje VoiceXML³. Este lenguaje facilita en gran medida la implementación de este tipo de sistemas, pues aísla la lógica de la aplicación (diálogo alumno-

³ <http://www.w3.org/TR/voicexml20/>

sistema) de las cuestiones de bajo nivel relacionadas con RAH, PLN, GLN o TTS.

La implementación se llevará a cabo usando un ciclo de vida de prototipado, en el cual se prevé desarrollar al menos dos prototipos hasta lograr un funcionamiento satisfactorio del sistema. En concreto, las tareas que realizaremos son las descritas a continuación.

Diseño

En esta tarea hemos realizado el diseño de las bases de datos que permitirán almacenar los materiales didácticos correspondientes a asignaturas de varias Titulaciones de la Universidad de Granada (preguntas y correspondientes respuestas). También hemos diseñado el modelo de la interacción sistema-alumno, como se muestra en la

Fig. 2. Asimismo, se ha comenzado a realizar el diseño del sistema de diálogo, lo cual implica crear los modelos de lenguaje para RAH, las reglas para PLN, las estrategias de interacción para GD, y los patrones para generación de respuestas del sistema en modo texto (GLN).

Implementación y evaluación de los prototipos

En esta tarea se implementará una primera versión del sistema de diálogo, es decir, un prototipo. Para implementar sus módulos de PLN y GLN, tomaremos como partida los módulos creados previamente por el equipo investigador para los sistemas SAPLEN (López-Cózar et al. 1999; López-Cózar et al., 2006) y VIAJERO (López-Cózar et al., 2000; López-Cózar y Callejas, 2005). Para implementar el módulo de GD, usaremos el lenguaje VoiceXML. Algunos scripts VoiceXML serán estáticos, mientras que otros serán generados dinámicamente mediante PHP. Para implementar el sistema de bases de datos, así como las consultas de forma remota al mismo, usaremos MySQL.

Una vez implementado el primer prototipo, se evaluará su funcionamiento para detectar posibles errores y corregirlos, obteniendo un prototipo mejorado. Ambas tareas (evaluación y corrección de errores) se repetirán cíclicamente hasta que el funcionamiento del prototipo se considere satisfactorio. La evaluación se realizará de forma subjetiva y objetiva. En el primer caso, utilizaremos cuestionarios anónimos mediante los cuales alumnos que se presten a probar el prototipo manifestarán sus opiniones acerca del funcionamiento del mismo, por ejemplo, en cuanto a la inteligibilidad y naturalidad de los mensajes generados mediante TTS, rapidez en la interacción, y facilidad para corregir errores cometidos por el prototipo. Estas cuestiones se realizarán online y se almacenarán en el ordenador servidor para su posterior análisis (ver Figura 1).

Para realizar la evaluación objetiva utilizaremos dos medidas típicamente usadas en la evaluación de sistemas de diálogo hablado: exactitud en el reconocimiento de palabras (word accuracy) y comprensión de frases (sentence understanding). La primera medida proporcionará información acerca

del funcionamiento del módulo de RAH, mientras que la segunda lo hará en relación al módulo de PLN.

Refinamiento y evolución de los prototipos

Con los resultados de la evaluación, se llevará a cabo un refinamiento del diseño en que se contemplen las mejoras necesarias para corregir los posibles errores detectados en los distintos módulos del prototipo. Quizás los cambios principales se deban realizar en las gramáticas para RAH, para tener en cuenta construcciones lingüísticas imprevistas, así como palabras no consideradas inicialmente en el diccionario del prototipo (

Figura 1.). Es posible que debamos realizar mejoras también en las estrategias de gestión del diálogo, por ejemplo, en lo concerniente a la iniciativa del diálogo y las estrategias de confirmación. Una vez se haya desarrollado un prototipo mejorado, se volverá a realizar su evaluación según el proceso descrito en la sección anterior, hasta que los resultados de evaluación sean satisfactorios.

Puesta en marcha del sistema para proporcionar el servicio pretendido

Una vez verificado el funcionamiento correcto del prototipo (de forma subjetiva y objetiva), se asumirá que éste constituye la versión final del sistema. En este momento, se dará a conocer a los alumnos de varias Titulaciones de la Universidad de Granada el número de teléfono al que podrán llamar para usar el sistema. Si lo desean, los alumnos podrán manifestar también sus opiniones acerca del funcionamiento del sistema mediante cuestionarios online, de forma similar a como lo hicieron los alumnos que colaboraron en la fase de implementación y evaluación de los prototipos.

CONCLUSIONES Y TRABAJO FUTURO

En este artículo hemos presentado una nueva herramienta basada en las nuevas tecnologías de la información y las comunicaciones, cuyo objetivo es favorecer el aprendizaje autónomo y la autoevaluación de los alumnos, de acuerdo con el nuevo Espacio Europeo de Educación Superior. En concreto, la herramienta es un sistema de diálogo hablado, llamado TRIVIAL, cuya finalidad será interactuar con alumnos a través del teléfono, con objeto de plantearles preguntas acerca de contenidos didácticos, obtener sus repuestas, indicarles si dichas respuestas son correctas o incorrectas, y en su caso, informales de las respuestas correctas. La interacción sistema-alumno se realizará mediante habla, por lo que el sistema se diferenciará en gran medida de las herramientas clásicas de autoaprendizaje y evaluación, basadas en el uso estático de la web. Dado que el proyecto de innovación docente en el que se va a desarrollar el sistema ha comenzado en el mes de enero de 2011, en este artículo nos hemos centrado en describir cuales son los objetivos del sistema, su arquitectura modular y las tareas que hemos desarrollado o tenemos previsto llevar a cabo. El trabajo futuro se centrará, pues, en la realización de las tareas pendientes.

BIBLIOGRAFÍA

- Ai, H., Litman, D. J., Forbes-Riley, K., Rotaru, M., Tetreault, J., Purandare, A. (2006). Using system and user performance features to improve emotion detection in spoken tutoring systems. *Actas de Interspeech*, págs. 797-800.
- Allen, J. F. (1995). *Natural Language Understanding*. Benjamin/Cummings.
- Gairín, J. Feixas, M., Guillamón, C., Quinquer, D. (2004). La tutoría académica en el escenario europeo de la Educación Superior. *Revista Interuniversitaria de Formación del Profesorado* 18, 1, págs. 61-77.
- García C. M., Lavié, J. M. (2000). Formación y nuevas tecnologías: Posibilidades y condiciones de la teleformación como espacio de aprendizaje. *Bordón* 52, 3, pp. 385-406.
- Griol, D., Hurtado, L. F., Segarra, E., Sanchis, E. (2008). A Dialogue Management Methodology Based on Neural Networks and Its Application to Different Domains. *LNCS 5197/2008*, págs. 643-650.
- Griol, D., García-Herrero, J., Callejas, Z., López-Cózar, Z. (2010). EducAgent: Agentes Virtuales para el Proceso de Enseñanza-Aprendizaje en Competencias. *Actas de ADA-MADRID II Jornadas Pedagógicas*, págs. 201-209.
- López-Cózar, R., Rubio, A. J., García, P., Segura, J. C. (1999). Uso de Valores de Confianza y Expectativas en el Sistema de Diálogo SAPLEN. *Procesamiento del Lenguaje Natural*, 24, págs. 37-41.
- López-Cózar, R., Rubio, A. J., García, P., Díaz-Verdejo, J. E., López-Soler, J. M. (2000). Sistema Telefónico de Atención a Viajeros. *Actas de I Jornadas en Tecnologías del Habla*.
- López-Cózar, R., Araki, M. (2005). *Spoken, Multilingual and Multimodal Dialogue Systems: Development and Assessment*. Ed. John Wiley & Sons.
- López-Cózar, R., Callejas, Z. (2005). Viajero III: Sistema de diálogo multimodal para proporcionar información de viajes en autobús. *Actas de Interacción'05*, págs. 329-336.
- López-Cózar, R., Callejas, Z., McTear, M. (2006). Testing the Performance of Spoken Dialogue Systems by Means of an Artificially Simulated User. *Artificial Intelligence Review*, 26, págs. 291-323.
- Rabiner, L. R., Huang, B. H. (1993). *Fundamentals of Speech Recognition*. Prentice-Hall.

Troyano, Y., García, A.J. (2007). Expectativas del alumnado sobre el profesorado tutor en el contexto del Espacio Europeo de Educación Superior. Red Estatal de Docencia Universitaria (REDU), seminario internacional 2-07: El desarrollo de la autonomía del aprendizaje.

Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T., Hetherington, L. 2000. Jupiter: A telephone-based conversational interface for weather information. IEEE Trans. on Speech and Audio Proc., 8(1), págs. 85-96.

Recibido: 11 marzo 2011.
Aceptado: 11 abril 2011.