Procesamiento del Lenguaje Natural, Revista nº 45, septiembre 2010, pp 311-312 recibido 01-05-10 revisado 18-05-10 aceptado 20-05-10

Kiosco de información basado en un sistema de diálogo multimodal*

A prototype based on a multimodal dialog system

J. Doncel J. M. Olaso R. Justo V. Guijarrubia A. Perez M. I. Torres
Universidad del País Vasco
manes.torres@ehu.es

Resumen: En este trabajo se describe un prototipo basado en un sistema de diálogo multimodal desarrollado utilizando la arquitectura EDECÁN para la empresa Softec-Ibermática. Duración de la demostración: 10 min. aprox.
Palabras clave: sistema de diálogo, multimodalidad

Abstract: The goal of this work is to develop a multimodal application based on a dialog system using EDECÁN architecture for Softec-Ibermática company. The demonstration takes 10 min. aprox.
Keywords: dialog system, multimodal interaction

1. Introduccion

El sistema de diálogo que se describe en este trabajo consiste en un kiosko cuyo objetivo es: proporcionar información meteorológica de diferentes ciudades del territorio español por un lado y por el otro, dar acceso a las noticias de las ediciones digitales de la prensa nacional. Además, se trata de un prototipo adecuado para el uso de personas con algún tipo de discapacidad debido a la posibilidad de interacturar con el sistema de forma multimodal.

Los dispositivos de entrada de este equipo son un teclado, un ratón, un micrófono y una web-cam y los dispositivos de salida son una pantalla y unos altavoces como puede verse en la Figura 1.

En este trabajo se presenta el conjunto de módulos que componen el proyecto. Estos módulos se han desarrollado de acuerdo con la arquitectura EDECÁN [1](García et al., 2008)(Lleida et al., 2006).

Figura 1: Sistema final

En la siguiente sección se describe cada uno de los módulos que conforman el sistema de diálogo que se describe en en este trabajo. Finalmente, se detallan las conclusiones más destacables así como posibles líneas de trabajo futuro a fin de mejorar el prototipo.

2. Arquitectura del sistema

Un sistema de diálogo como el que se ilustra en la Figura 2 considera las interacciones del usuario como entrada y proporciona al usuario una salida que generalmente consiste en acciones o instrucciones orales generadas de forma sintética.

A continuación, se detallan cada uno de los módulos del sistema implementados bajo la arquitectura EDECÁN:

* Este trabajo ha sido parcialmente financiado por la Consejería de Industria del Gobierno Vasco, la Universidad del País Vasco en el proyecto GIU07/57 y por el Ministerio de Ciencia e Innovación mediante el proyecto TIN2008-06856-C05-01 y por el programa Consolider Ingenio-2010 MIPRCV (CSD2007-00018)

[1]Proyecto coordinado financiado por el Ministerio de Ciencia e Innovación, 2005-2008, liderado por la Universidad de Zaragoza y con la participación de la Universidad del País Vasco, la Univeridad Politécnica de Valencia y la Universidad Politécnica de Madrid. http://www.edecan.es

ISSN 1135-5948 © 2010 Sociedad Española para el Procesamiento del Lenguaje Natural

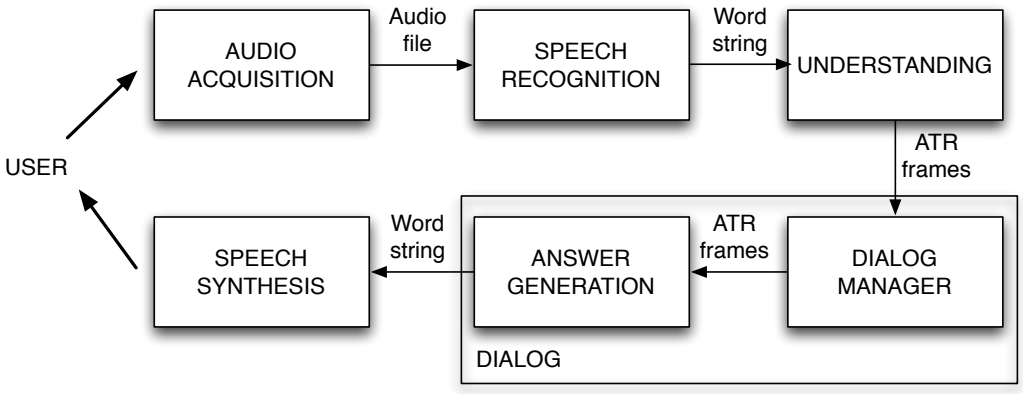

Figura 2: Arquitectura básica de un sistema de diálogo

- Módulo de detección de caras: se trata del módulo que activa el sistema, ya que mientras no se detecte la presencia del usuario el sistema permanece inactivo.
- Módulo de reconocedor de voz (ASR): se trata del módulo que proporciona la secuencia de palabras que con mayor probabilidad se ajusta a la señal acústica de entrada. El reconocedor implementado en este módulo ha sido desarrollado integramente por el grupo de Reconocimiento de Formas y Tecnologías del habla de la Universidad del País Vasco. Se trata de un reconocedor que integra modelos acústicos implementados mediante modelos ocultos de Markov contínuos y modelos de lenguaje implementados mediante autómatas de estados finitos estocásticos.
- Módulo de comprensión: se trata del módulo que recoge las palabras extraidas por el reconocedor de voz y se encarga de extraer la información conceptual de la secuencia de palabras obtenida. Para realizar esta función, se transforma la secuencia de palabras en unidades semánticas que contienen información sobre la tarea a realizar.
- Módulo gestor de diálogo: a partir de la información semántica proporcionada, decide la estrategia a seguir por el proceso y va dando paso al sistema y al usuario.
- Módulo cliente-web: la informacón requerida por el usuario se obtiene de internet mediante el módulo cienteweb. Para realizar esta función se descarga la página solicitada en formato xml.
- Módulo xlst: este módulo convierte la información requerida de formato xml a html, que es el formato adecuado para mostrarlo por pantalla.
- Módulo visualización: este módulo dispone de un avatar y la imagen que obtiene la web-cam. Para generar el avatar se ha usado la herramienta *iclone*, la versión libre de *Reallusion*. La imagen de la web-cam permite conocer el estado del detector de caras.
- Módulo conversor texto a voz (TTS): este módulo tiene como objetivo generar artificialmente una señal acústica, a partir de una secuencia de palabras de entrada. Para llevar a cabo esta tarea se ha empleado el sistema Loquendo TTS, asi como unos altavoces que permite al usuario oír la señal acústica de salida.

3. Conclusiones

En el diseño del sistema de diálogo descrito en este trabajo se dispone de una interacción multimodal entre usuario y máquina. De este modo, se consigue superar las limitaciones que un sistema de diálogo basado únicamente en habla impone.

En cuanto a las líneas de mejora para un futuro trabajo se encuentran el diseño de un prototipo que permita interacción via web, de tal modo que de cualquier ordenador se pueda acceder a esta información.