

ADQUISICIÓN DE UN CORPUS DE DIÁLOGOS PARA UN DOMINIO DE RESERVAS DE INSTALACIONES DEPORTIVAS

*E. Segarra, M.J. Castro, I. Galiano, F. García, J. A. Gómez, D. Griol,
L.F. Hurtado, E. Sanchis, F. Torres, F. Zamora*

Departament de Sistemes Informàtics i Computació
Universitat Politècnica de València, 46022 València, Spain

`esegarra@dsic.upv.es`

RESUMEN

La adquisición de un corpus de diálogos hablados es un proceso complejo y costoso. Con el objetivo de facilitar este proceso proponemos un método para llevar a cabo la adquisición de un corpus de diálogos para una tarea de requerimientos a un sistema de información; en particular, en este trabajo el usuario interacciona con un sistema de información de reservas de pistas deportivas. En este trabajo presentamos una descripción de los diferentes componentes del sistema de adquisición.

1. INTRODUCCIÓN

El desarrollo de sistemas de diálogo hablado ha recibido un fuerte impulso en los últimos años en el marco de las tecnologías del lenguaje hablado. El desarrollo de estos sistemas es un proceso complejo que comporta el diseño, implementación y evaluación de varios módulos en los que intervienen diversas fuentes de conocimiento. Actualmente, una de las aproximaciones adoptadas con más éxito es la basada en modelos estadísticos, que representan de forma probabilística los diferentes procesos implicados en los módulos y en la que los correspondientes modelos se estiman a partir de corpus de diálogos persona-máquina [1, 2, 3, 4]. El éxito de las aproximaciones estadísticas depende principalmente de la calidad de los modelos y, por tanto, de los corpus a partir de los cuales son entrenados. Es por ello que la adquisición de corpus adecuados y la definición de una adecuada representación semántica para su etiquetado son procesos claves.

La adquisición de un corpus puede resultar un proceso complejo y costoso por lo que es útil cualquier ayuda que permita facilitar este proceso. Con este objetivo en la adquisición de un corpus de diálogos para la tarea EDECAN-SPORT de reservas de pistas deportivas en el marco del proyecto Edecán [5] hemos seguido el proceso que describimos a continuación.

En primer lugar, hemos analizado un conjunto de diálogos persona-persona facilitados por el servicio de reservas de pistas deportivas de nuestra universidad, que constituye el dominio de trabajo de la tarea EDECAN-SPORT. A partir de estos diálogos se ha definido la semántica de la tarea en términos de frames de los turnos de usuario y de los turnos de sistema, y se han etiquetado estos diálogos iniciales. Con ello disponemos de un corpus inicial muy reducido para la tarea EDECAN-SPORT. A partir de este pequeño corpus etiquetado se ha obtenido una versión preliminar de un gestor de diálogo que se estima a partir de datos [3]. Este gestor de diálogo se utiliza en el proceso de adquisición de un corpus más amplio de esta tarea con la ayuda de la técnica del Mago de Oz (las adquisiciones se llevan a cabo con usuarios reales y un sistema de diálogo simulado en parte).

En segundo lugar, como el corpus inicial no tiene la suficiente tamaño para entrenar modelos adecuados para el módulo de comprensión del habla, no disponemos de una versión preliminar de este módulo para el proceso de adquisición con el Mago de Oz. Nuestra propuesta consiste en el uso de un Mago de Oz específico para el módulo de comprensión diferente del Mago de Oz que supervisa el gestor de diálogo. La idea de usar más de un Mago de Oz fue propuesta por [6].

Para que este Mago específico de comprensión actúe de la forma más parecida a un sistema automático, va acompañado de un proceso que realiza una simulación de errores de reconocimiento y comprensión. Esta simulación [7] está basada en el análisis de los errores de reconocimiento y comprensión generados cuando nuestros módulos de reconocimiento y comprensión son aplicados a un corpus adquirido en anteriores proyectos [8], también para una tarea de consulta a un sistema de información.

En este trabajo presentamos una descripción de los diferentes componentes del sistema de adquisición para la tarea EDECAN-SPORT. En la sección 2 describimos el corpus inicial, y en la sección 3 la representación semántica tanto de los turnos de usuario como de sistema. En la sección 4 describimos la arquitectura global del sistema de adquisición y en la sección 5 los datos globales del corpus que se está adquiriendo.

Trabajo parcialmente subvencionado por el gobierno español y los fondos FEDER con el proyecto TIN2005-08660-C04-02, y por el Programa de Apoyo a la Investigación y Desarrollo PAID-05-08 de la Universitat Politècnica de València.

S1: Bienvenido al servicio de reservas de instalaciones deportivas. ¿En qué puedo ayudarle?
 U1: Quiero reservar una pista de pádel para el próximo miércoles.
 S2: ¿A qué hora quiere jugar?
 U2: A las tres de la tarde.
 S3: El día 8 de diciembre a las 15:00 horas están disponibles las pistas de pádel número 1 y 4. ¿Cuál quiere reservar?
 U3: La pista de pádel número 1.
 S4: Le acabo de reservar la pista de pádel número 1 para el día 8 de diciembre a las 15:00 horas. ¿Quiere algo más?
 U4: Sí, quiero cancelar una pista para mañana por la tarde.
 S5: ¿Quiere cancelar la reserva de la pista de tenis número 3 del día 7 de diciembre a las 19:00?
 U5: Sí, por favor.
 S6: Acabo de cancelar la pista señalada. ¿Quiere algo más?
 U6: No, gracias.
 S7: Gracias por usar este servicio de reservas. Adiós.

Figura 1. Un ejemplo de diálogo para la tarea EDECAN-SPORT.

2. DIÁLOGOS PERSONA-PERSONA

Se dispone de un pequeño conjunto de diálogos persona-persona reales de la aplicación de reservas de pistas deportivas, gracias a la colaboración del personal del servicio de deportes de la universidad. Estos diálogos han facilitado la definición del dominio semántico de la tarea EDECAN-SPORT: consultas de horarios disponibles, consultas de reservas realizadas anteriormente, peticiones de reservas de pistas deportivas y, finalmente, peticiones de cancelaciones de reservas anteriores. Se ha obtenido un total de 150 diálogos que comprenden 837 turnos de usuario. La Figura 1 muestra un ejemplo de estos diálogos.

3. DEFINICIÓN DE LA SEMÁNTICA DE LA TAREA EDECAN-SPORT

La definición de la semántica de la tarea se llevó a cabo a partir de los anteriores diálogos y teniendo en cuenta las diferentes prestaciones del servicio real ofrecido por la universidad.

3.1. Representación semántica de los turnos de usuario

Los actos de diálogo de usuario representan la interpretación semántica de los turnos de usuario en términos de frames (conceptos y atributos). Para la tarea EDECAN-SPORT hemos identificado seis conceptos:

- Cuatro conceptos dependientes de la tarea y relacionados con los diferentes tipos de interacción del usuario con el sistema (requerimientos al sistema de información): disponibilidad de pistas *Availability*, reserva de pistas *Booking*, consulta de pistas reservadas *Booked*, y cancelación de reservas *Cancellation*.
- Dos conceptos independientes de la tarea: *Acceptance* y *Rejection*.

Han sido identificados un total de seis atributos. Están relacionados con la información que el usuario debe proporcionar al sistema para completar los requerimientos al

sistema de información. Estos atributos son: *Sport*, *Hour*, *Date*, *Court-Type*, *Court-Number*, and *Order-Number*.

A continuación se muestra un ejemplo de la representación semántica de un turno de usuario:

Turno de usuario: *Quiero reservar una pista de tenis para mañana por la tarde*

Representación semántica: (Booking)

Sport: tenis
 Date: mañana
 Hour: tarde

3.2. Representación semántica de los turnos de sistema

Los turnos de sistema han sido también etiquetados. Los conceptos se han dividido también en dependientes e independientes de la tarea. Se ha definido un conjunto de 15 conceptos dependientes de la tarea:

- Cuatro conceptos principales que informan al usuario acerca del resultado de un requerimiento: acerca de la disponibilidad de pistas *Availability*, acerca de la reserva de pistas *Booking*, acerca de las reservas actuales *Booked* y acerca de la cancelación de reservas *Cancellation*.
- Cuatro conceptos para pedir al usuario los valores de atributos necesarios para llevar a cabo la consulta al sistema de información: *Sport*, *Date*, *Hour*, and *Court-Type*.
- Cuatro conceptos para confirmar los cuatro conceptos principales y un concepto *Confirmation* para confirmar los diferentes valores de atributos.
- Otros conceptos: violación de las normas de reserva del servicio de la universidad *Rule-Info* y la petición de selección de una de las pistas disponibles *Booking-Choice*.

Han sido identificados seis atributos: cinco relacionados con atributos del turno de usuario (*Sport*, *Court-Type*, *Court-Number*, *Date*, *Hour*), y un atributo relacionado

con el número de pistas que satisfacen el requerimiento del usuario (*Availability-Number*). Por otra parte, ha sido identificado un conjunto de seis conceptos independientes de la tarea: *Opening*, *Closing*, *Non-Understood*, *New-Query*, *Acceptance* y *Rejection*. A continuación se presenta un ejemplo de etiquetado de turno de sistema:

Turno de sistema: *¿Quiere reservar la pista de squash número 1 en el pabellón para el 25 de junio de 20:00 a 20:30 horas?*

Representación semántica: (Confirmation-Booking)

Sport: squash
Date: 25-06-2008
Hour: 20:00-20:30
Court-Type: pabellón
Court-Number: 1

4. LA ARQUITECTURA PARA LA ADQUISICIÓN DEL CORPUS PARA LA TAREA EDECAN-SPORT

Siguiendo las principales contribuciones de la literatura en el área de los sistemas de diálogo hablado, realizamos la adquisición del corpus para la tarea EDECAN-SPORT usando la técnica del Mago de Oz. En esta técnica, una persona sustituye al sistema de diálogo en la mayor parte de sus funciones, principalmente: escucha al locutor, construye el requerimiento al sistema de información y construye el frame de sistema como respuesta al locutor. Un generador de lenguaje natural construye la respuesta al locutor en base a este frame. Finalmente, se genera el audio correspondiente a esta frase en lenguaje natural.

Usualmente estas tareas son realizadas por una única persona. Nuestra propuesta [7] para esta adquisición estriba en trabajar con dos Magos de Oz: un simulador del proceso de comprensión del habla y un simulador del gestor de diálogo. El primero escucha al locutor y simula la acción de los módulos de reconocimiento y comprensión del habla y proporciona la representación semántica del turno de usuario (el frame/s correspondiente/s). A partir de dicho frame, el segundo Mago actúa como un gestor de diálogo. La arquitectura que se propone se muestra en la Figura 2.

Además de simplificar las funciones del Mago de Oz, la separación de la comprensión y del gestor de diálogo permite simular con más fidelidad lo que ocurre en un sistema totalmente automático. El gestor de diálogo no escucha al locutor, sino que recibe la información proporcionada únicamente por el módulo de comprensión. Con ello conseguimos similares condiciones experimentales en el proceso de adquisición que en el proceso de evaluación y prueba del sistema de diálogo.

4.1. El simulador de comprensión

El Mago de Oz de comprensión, utilizando el llamado editor de comprensión, traduce la intervención del locutor en su correspondiente frame (que nos va a servir co-

mo frame de referencia en el etiquetado de los turnos de usuario del corpus). A partir de este frame, el simulador de errores introduce ciertos errores en el frame de referencia generando el frame simulado, que será el que se proporcionará al segundo Mago. Este proceso se muestra en la Figura 3. El simulador de errores reproduce el comportamiento de nuestros módulos de reconocimiento y comprensión desarrollados en el marco del proyecto DIHANA [8].

4.2. El simulador del gestor de diálogo

Hemos desarrollado una aproximación para la gestión del diálogo haciendo uso de un modelo estadístico que se estima a partir de corpus [3] en el marco del proyecto DIHANA [8]. En esta aproximación, la respuesta del gestor en un momento del diálogo se selecciona en base a un proceso de clasificación que tiene en cuenta la información suministrada por el usuario a lo largo del diálogo y la última intervención del mismo. Este modelo de gestor se estima automáticamente a partir de un corpus de diálogos de entrenamiento etiquetado en términos de actos de diálogo. Este gestor almacena la información proporcionada por el usuario en una estructura de datos que contiene también medidas de confianza para cada uno de los elementos de esta estructura (conceptos y atributos).

Se ha adaptado esta aproximación para su uso en el marco del proyecto EDECÁN, en particular para la tarea de reservas de instalaciones deportivas EDECÁN-SPORT [9]. Esta adaptación tiene en cuenta los nuevos requerimientos introducidos por esta tarea, lo que incluye el uso de un manejador de aplicación que interactúa con la base de datos y que verifica si el requerimiento del usuario cumple con las normas definidas para el uso del servicio de reservas. Como ya se ha comentado en la introducción, a partir de los diálogos persona-persona etiquetados en términos de frames se ha estimado una versión preliminar de gestor de diálogo para la tarea EDECÁN-SPORT.

Esta versión está implementada en el sistema de adquisición, y la misión del Mago de Oz de diálogo consiste en supervisar su funcionamiento. Esta supervisión se lleva a cabo a través de dos aplicaciones, en una se supervisa (se corrige cuando el sistema automático ha fallado) el funcionamiento del gestor de diálogo y en la otra aplicación se supervisa el funcionamiento del manejador de la aplicación.

5. LA ADQUISICIÓN

Se están adquiriendo 240 diálogos en los que intervienen 18 locutores (9 hombres y 9 mujeres) de diferentes procedencias geográficas (las 4 sedes de los equipos investigadores del consorcio EDECÁN). Las lenguas implicadas en la adquisición son tres: castellano, catalán y eusquera. Se ha definido un conjunto de 15 tipos de escenarios. A continuación se muestra un escenario.

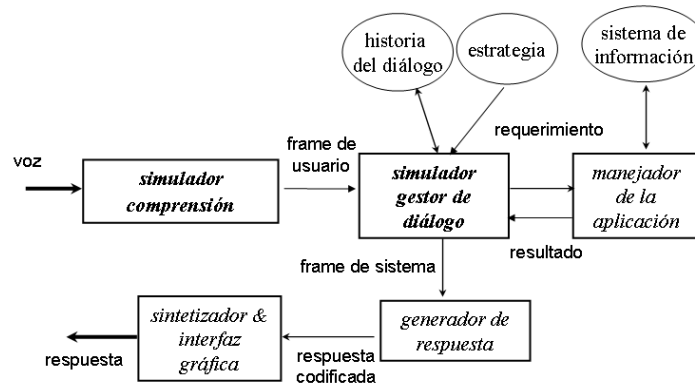


Figura 2. Esquema de la adquisición.

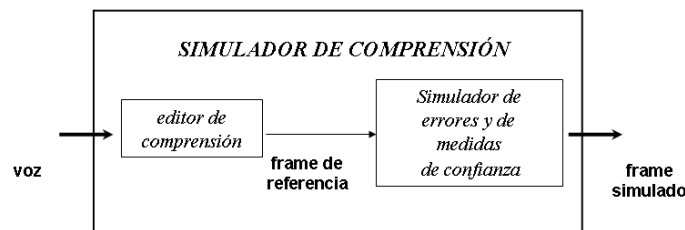


Figura 3. Simulador de comprensión.

Objective: Conocer la disponibilidad y reservar.
Sport: Padel.
Date: miércoles, jueves o viernes.
Hour: tarde.

6. BIBLIOGRAFÍA

- [1] A. Potamianos, E. Ammicht, y E. Fosler-Lussier, “Modality tracking in the Multimodal Bell Labs Communicator,” in *Proc. of ASRU 03*, St. Thomas, U.S. Virgin Islands, 2003, pp. 192–197.
- [2] F. Torres, L.F. Hurtado, F. García, E. Sanchis, y E. Segarra, “Error handling in a stochastic dialog system through confidence measures,” *Speech Communication*, vol. 45, pp. 211–229, 2005.
- [3] L.F. Hurtado, D. Griol, E. Segarra, y E. Sanchis, “A Stochastic Approach for Dialog Management based on Neural Networks,” in *Procs. of InterSpeech’06*, Pittsburgh, USA, 2006, pp. 49–52.
- [4] J. Williams y S. Young, “Partially Observable Markov Decision Processes for Spoken Dialog Systems,” in *Computer Speech and Language* 21(2), 2007, pp. 393–422.
- [5] Eduardo Lleida, Encarna Segarra, M. Inés Torres, y J. Macías-Guarasa, “EDECAN: sistEma de Diálogo multidominio con adaptación al contExto aCústico y de Aplicación,” in *IV Jornadas en Tecnología del Habla*, Zaragoza, Spain, 2006, pp. 291–296.
- [6] Daniel Salber y Joëlle Coutaz, “Applying the wizard of oz technique to the study of multimodal systems,” in *EWHCI ’93: 3rd International Conference on Human-Computer Interaction*, London, UK, 1993, pp. 219–230, Springer-Verlag.
- [7] F. Garcia, L.F. Hurtado, D. Griol, M. Castro, E. Segarra, y E. Sanchis, “Recognition and Understanding Simulation for a Spoken Dialog Corpus Acquisition,” in *TSD 2007*, vol. 4629 of *LNAI*, pp. 574–581. Springer, 2007.
- [8] José-Miguel Benedí, Eduardo Lleida, Amparo Varona, María-José Castro, Isabel Galiano, Raquel Justo, Iñigo López de Letona, y Antonio Miguel, “Design and acquisition of a telephone spontaneous speech dialogue corpus in Spanish: DIHANA,” in *Proceedings of LREC 2006*, Genoa (Italy), May 2006, pp. 1636–1639.
- [9] Lluís F. Hurtado, David Griol, Encarna Segarra, y Emilio Sanchis, “Adapting a Statistical Dialog for a New Domain,” in *Proc. of DECALOG’07*, Rovereto (Italy), 2007, pp. 171–172.