



ESTUDIOS / RESEARCH STUDIES

CALIMACO: desarrollo de un servicio de bibliotecario virtual para la interacción multimodal con dispositivos móviles

David Griol*, Miguel Ángel Patricio*, José Manuel Molina*

* Grupo de Inteligencia Artificial Aplicada (GIAA), Departamento de Ciencias de la Computación. Universidad Carlos III de Madrid.

Correos-e: david.griol@uc3m.es, miguelangel.patricio@uc3m.es, josemanuel.molina@uc3m.es

Recibido: 07-02-2015; 2ª versión: 20-07-2015; Aceptado: 25-07-2015.

Cómo citar este artículo/Citation: Griol, D.; Patricio, M. A.; Molina, J. M. (2016). CALIMACO: desarrollo de un servicio de bibliotecario virtual para la interacción multimodal con dispositivos móviles. *Revista Española de Documentación Científica*, 39(2): e129. doi: <http://dx.doi.org/10.3989/redc.2016.2.1262>

Resumen: Los dispositivos móviles inteligentes han fomentado en los últimos años nuevos escenarios de interacción que requieren interfaces persona-máquina más sofisticados. Los principales desarrolladores de sistemas operativos para estos dispositivos proporcionan actualmente APIs (*Application Programming Interfaces*) para que los desarrolladores puedan implementar sus propias aplicaciones, incluyendo diferentes soluciones para el desarrollo de interfaces gráficas, el control de sensores y proveer interacción oral. A pesar de la utilidad de este tipo de recursos, no se han definido aún estrategias para desarrollar interfaces multimodales que aprovechen las posibilidades que ofrecen estos dispositivos para identificar y adaptarse a las necesidades de los usuarios. De este modo, las aplicaciones suelen desarrollarse ad-hoc y la interfaz oral posibilita únicamente la comunicación mediante comandos simples. En este artículo proponemos la aplicación práctica de agentes conversacionales multimodales sensibles al contexto para proporcionar servicios bibliotecarios avanzados que incorporen dinámicamente necesidades específicas del usuario y sus preferencias, así como las características específicas del entorno en el que se produce interacción, con el fin de mejorar y personalizar el servicio que se proporciona mediante un dispositivo móvil con acceso a Internet. Nuestra propuesta integra funcionalidades de APIs de Android en una arquitectura modular que enfatiza la gestión de la interacción y la sensibilidad al contexto para crear aplicaciones robustas, de fácil actualización y adaptadas al usuario.

Palabras clave: Sistemas de Diálogo; interacción multimodal; reconocimiento y síntesis de voz; asistente; servicios bibliotecarios avanzados; dispositivos móviles; Android.

CALIMACO: Application of multimodal dialog systems and mobile devices to provide enhanced library services

Abstract: In recent years smart mobile devices have bolstered new interaction scenarios that require more sophisticated human-machine interfaces. The leading developers of operating systems for these devices now provide APIs (*Application Programming Interface*) for developers to implement their own applications, including different solutions for developing graphical interfaces, control sensors and providing oral interaction. Despite the usefulness of these resources, defined strategies are still needed for developing multimodal interfaces to take greater advantage of these devices for identifying and meeting the needs of users. Currently, these applications are typically ad-hoc and facilitate oral communication only through simple commands. In this paper we propose the practical application of context-sensitive multimodal conversational agents to provide advanced library services that dynamically consider specific user needs and preferences, as well as the specific characteristics of the environment in which the interaction occurs. Such agents would improve and customize the service provided by a mobile device with Internet access. Our proposal integrates features of Android APIs on a modular architecture emphasizing the management of interactions and context awareness in order to create robust applications that can be easily updated and adapted to the user.

Keywords: Dialog systems; multimodal interaction; automatic speech recognition and synthesis; automatic assistants; advanced library services; mobile devices; Android.

Copyright: © 2016 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Attribution-Non Commercial (by-nc) Spain 3.0.

1. INTRODUCCIÓN

El incremento de uso de los dispositivos móviles, tales como los teléfonos inteligentes (*smartphones*) y tablet-PCs, ha propiciado que un número continuamente creciente de personas lleven siempre consigo un dispositivo con numerosas modalidades de presentación de la información, sensores y posibilidades de conectividad. Ello ha generado nuevos escenarios de interacción que requieren interfaces persona-máquina más sofisticadas (McTear y Callejas, 2013; Becker y otros, 2013).

Los sistemas de diálogo multimodales van un paso más allá de las interfaces gráficas de usuario tradicionales (GUI), posibilitando la comunicación con estos dispositivos a través de otros modos de interacción, como el habla, de forma táctil y/o visual (McTear y Callejas, 2013; Pieraccini, 2012; Pérez-Marín y Pascual-Nieto, 2011).

Estos sistemas son programas de ordenador desarrollados para emular las capacidades de comunicación de los seres humanos mediante diferentes modalidades, con la finalidad de proporcionarles información o servicios automatizados de una forma más natural. La interacción se lleva a cabo mediante turnos de un tipo de diálogo que los investigadores intentan que se parezca lo más posible al diálogo real que se lleva a cabo entre las personas, en lo que se refiere a naturalidad, inteligencia e incluso contenido afectivo.

Sin embargo, los sistemas de diálogo se diseñan generalmente ad-hoc para su dominio específico de aplicación, utilizando modelos y estándares en los que los desarrolladores deben especificar las reglas a seguir que debe seguir el sistema tras cada intervención del usuario (Larson, 2002). De este modo, la adaptación del sistema teniendo en cuenta las preferencias específicas de los usuarios o la aplicación del sistema a nuevos dominios de interacción es un proceso que requiere un tiempo y esfuerzo considerables (Paek y Pieraccini, 2008).

Además, los dispositivos móviles han originado un nuevo paradigma que posibilita recopilar información del usuario de modo no intrusivo. Ello permite construir modelos de usuario más complejos que pueden emplearse no sólo para proporcionar la funcionalidad específica del sistema, sino también para mejorar su rendimiento. Sin embargo, esta información no se suele considerar al diseñar el modelo de diálogo del sistema (Seneff y otros, 2007). Por esta razón, en la mayoría de sistemas de diálogo, la respuesta proporcionada por el sistema en cada uno de los casos definidos es la misma para todos los usuarios, de forma que éstos normalmente no tienen ningún control sobre el contenido o la presentación del servicio prestado.

Para resolver este problema, es necesario diseñar cuidadosamente tanto el modelo de usuario como la estrategia de gestión del diálogo. La aplicación de metodologías estadísticas para lograr este propósito posibilita que ambos modelos puedan entrenarse a partir de diálogos reales, modelando la variabilidad de comportamientos de los usuarios (Griol y otros, 2014). Aunque la parametrización de los modelos estadísticos depende del conocimiento experto de la tarea, los sistemas de diálogo resultantes tienen un comportamiento más robusto, una mayor portabilidad, y son más fáciles de adaptarse a diferentes perfiles de usuario o tareas (Schatzmann y otros, 2006).

En este trabajo, proponemos una arquitectura para el desarrollo de sistemas multimodales adaptados al contexto de la interacción y orientados a su funcionamiento en dispositivos móviles. Nuestra propuesta se basa en la definición de una metodología estadística para el modelado de usuario que estima la intención del usuario durante el diálogo, modelando ésta el siguiente paso que debe llevar a cabo el usuario tras cada una de las intervenciones del sistema.

En nuestra propuesta, la información proporcionada por el modelo de usuario se enriquece al considerar también información correspondiente al contexto externo de la interacción. Esta información se adquiere por medio de los sensores integrados en los dispositivos móviles. Actualmente el 75% de los teléfonos inteligentes y las tablet-PCs integran el sistema operativo Android (McTear y Callejas, 2013). Además, existe una comunidad activa de desarrolladores que utilizan el proyecto Android Open Source y que han desarrollado ya más de un millón de aplicaciones disponibles actualmente en la Play Store oficial, muchas de ellas totalmente gratuitas. Por estas razones, nuestra propuesta utiliza diferentes funcionalidades integradas actualmente en los dispositivos móviles Android.

El contexto interno y externo, respectivamente, en relación con la predicción de la intención del usuario por el modelo de usuario y la información proporcionada por los sensores, permiten adaptar el sistema de forma dinámica teniendo en cuenta estas valiosas fuentes de información. Para ello, se utiliza un modelo de diálogo estadístico basado en redes neuronales, que tiene en cuenta la información contextual y la historia de diálogo hasta el momento actual para decidir la siguiente respuesta del sistema. La codificación de la información y la definición de una estructura de datos que tiene en cuenta la historia del diálogo posibilitan el aprendizaje del modelo de diálogo de los datos de entrenamiento y dominios prácticos y manejables.

El resto del artículo se estructura de la siguiente forma. La Sección 2 describe brevemente la motivación de nuestra propuesta y el estado del arte relacionado con la misma. La Sección 3 describe la arquitectura propuesta para desarrollar interfaces multimodales adaptados al contexto para interfaces móviles. La Sección 4 describe la aplicación práctica de nuestra propuesta para el desarrollo de una aplicación multimodal sensible al contexto que proporciona servicios bibliotecarios avanzados que incorporan dinámicamente necesidades específicas del usuario y sus preferencias, así como características propias del entorno en el que están interactuando con la aplicación. La Sección 5 muestra los resultados de la evaluación de dicho sistema con usuarios reales. Finalmente, la Sección 6 presenta las conclusiones del trabajo y las líneas de trabajo futuro.

2. ESTADO DEL ARTE

La información digital permite el almacenamiento, acceso y transmisión de millones de recursos de forma sencilla, casi instantánea y desde prácticamente cualquier lugar gracias a los dispositivos móviles. Las bibliotecas digitales constituyen un ejemplo muy importante de estos recursos, pudiéndose definir como una extensión de las bibliotecas clásicas donde la información sobre las diferentes temáticas puede consultarse a través de la web (Balagué y otros, 2014; Ross y Sennyey, 2008). La aparición de estas bibliotecas ha cambiado la percepción tradicional de las mismas (Díaz-Valenzuela y otros, 2013; Liu, 2006).

En nuestro trabajo estamos especialmente interesados en las Bibliotecas Digitales Universitarias (*University Digital Libraries*, UDL) (Vassilakaki y Garoufallou, 2013). Este tipo de bibliotecas almacenan información sobre libros, artículos científicos en revistas y congresos, diarios oficiales, y usuarios registrados (Renda y Straccia, 2005; Smeaton y Callan, 2005; Calhoun, 2002). Los avances tecnológicos asociados con la denominada Web 2.0 (O'Reilly, 2005) han posibilitado la aparición de un gran número de nuevos paradigmas y servicios relacionados con las UDL, en los que el usuario es el actor principal y puede crear, modificar y publicar de forma sencilla nuevos contenidos web en colaboración con otros usuarios a través de nuevas aplicaciones y dispositivos como los blogs, wikis, chats, podcasts, gestores RSS (formato para compartir contenido en la web, abreviado del inglés: *Really Simple Syndication*), las redes sociales o los nuevos dispositivos móviles inteligentes (Fagan, 2014). Ello ha originado la aparición del término Bibliotecas 2.0 (*Library 2.0*) (Casey y Savastinuk, 2007).

La aplicación de estas tecnologías y uso de estos nuevos dispositivos es uno de los objetivos de nuestro trabajo para extender el concepto de Biblioteca 2.0. Para ello, nos hemos basado en las recomendaciones para el desarrollo de las bibliotecas de la era Web recopiladas por estudios y trabajos tan importantes como (Orduña-Malea, 2013; Chen y Albee, 2012), (Calhoun y Cellentani, 2009; Mann, 2008) en los que se proponen aspectos como:

- Incrementar la eficiencia en la producción de recursos bibliográficos mediante la cooperación entre las bibliotecas a través de la aplicación de las nuevas tecnologías, posicionando a las bibliotecas a través del reconocimiento de las posibilidades que ofrece la WWW (World Wide Web).
- Expandir los servicios y número de usos posibles de las bibliotecas mediante la incorporación de nuevos contenidos y tipos de accesos a los mismos.
- Adaptar la provisión de los contenidos y servicios de acuerdo a los requisitos, preferencias y perfiles de usuario.
- Incrementar el número de usuarios y el grado de satisfacción de los mismos a través de las nuevas tecnologías y servicios asociados.
- Facilitar la supervivencia de las bibliotecas incrementando los beneficios técnicos, de personal y financieros para su personal y usuarios.

Un gran número de trabajos han aplicado estas recomendaciones basándose en aspectos tan importantes como el uso de programas y aplicaciones de código y acceso libre para la colaboración y creación de consorcios bibliotecarios con un bajo coste económico (Park y otros, 2009; Chen y Albee, 2012; Helling, 2010; Evergreen, 2010); la utilización de dispositivos RFID (*Radio Frequency Identification*) para facilitar los servicios de inventariado, catalogación y préstamo (Moreno y otros, 2014; Dwivedi y otros, 2013; Sanz-Santacruz y otros, 2013); la utilización de tecnologías de realidad virtual y realidad aumentada, conjuntamente con los nuevos dispositivos móviles, para el desarrollo de aplicaciones bibliotecarias inmersivas (Shatte y otros, 2014; Chen y Tsai, 2012; González-Fernández-Villavicencio, 2014); o la incorporación de los sistemas de recomendación para proveer servicios bibliotecarios adaptados a los requerimientos y preferencias de los usuarios (Tejeda-Lorente, y otros, 2014; Tsuji y otros, 2014; Porcel y Herrera-Viedma, 2010; Porcel y otros, 2009; Renda y Straccia, 2005; Serrano-Guerrero y otros, 2011).

En el caso especial de aplicación de nuevas tecnologías para el apoyo de los servicios de una biblioteca destaca el trabajo realizado por Castaño

y otros, 2013. En el trabajo se utiliza tecnología RFID para la localización de los libros y la orientación en el edificio de la Biblioteca Municipal de Meco (Madrid). En línea con estos sistemas, y tal y como se ha descrito en la introducción del artículo, nuestro trabajo se centra en la aplicación de los sistemas de diálogo multimodal para la consecución de las recomendaciones descritas para proporcionar servicios bibliotecarios avanzados. El diccionario de la Real Academia Española de la Lengua define el diálogo como la plática entre dos o más personas que alternativamente muestran sus ideas o afectos. La acción de conversar es el modo más natural para resolver un gran número de acciones cotidianas entre los seres humanos: obtener una determinada información, contratar un servicio, solicitar un pedido, conocer el estado de un determinado proceso, etc.

Por este motivo, un interés histórico dentro del campo de las Tecnologías del Habla ha sido utilizar estas tecnologías en aplicaciones reales, especialmente en aplicaciones que permitan a una persona utilizar su voz para obtener información mediante la interacción directa con una máquina o para controlar un determinado sistema. El objetivo es disponer de sistemas que faciliten la comunicación persona-máquina del modo más natural posible, es decir, a través de la conversación.

Construir una aplicación informática que pueda mantener una conversación con una persona de manera natural sigue siendo hoy en día un reto, dada la gran cantidad de fuentes de conocimiento que son necesarias y las limitaciones de las tecnologías utilizadas para obtener información del usuario. No obstante, los constantes avances de la investigación en Tecnologías del Habla han permitido que sean factibles actualmente sistemas de comunicación persona-máquina mediante la voz, capaces de interactuar con cierto grado de flexibilidad. A nuestro entender, no existen trabajos previos que hayan aplicado el desarrollo de sistema de diálogo multimodal para la provisión de servicios bibliotecarios, posibilitando además que estos servicios sean accesibles de forma personalizada para cada uno de los usuarios y mediante los requisitos y posibilidades que ofrecen actualmente los dispositivos móviles avanzados.

3. PROPUESTA PARA EL DESARROLLO DE SISTEMAS DE DIÁLOGO MULTIMODAL PARA DISPOSITIVOS MÓVILES ANDROID

La Figura 1 muestra la arquitectura propuesta para el desarrollo de sistemas de diálogo multimodales adaptados al usuario y orientados a dispositivos móviles. Un sistema de diálogo hablado in-

tegra cinco módulos fundamentales: Reconocedor Automático del Habla, Módulo de Comprensión del Habla, Gestor de Diálogo, Generador de Lenguaje Natural y Sintetizador de Texto a Voz.

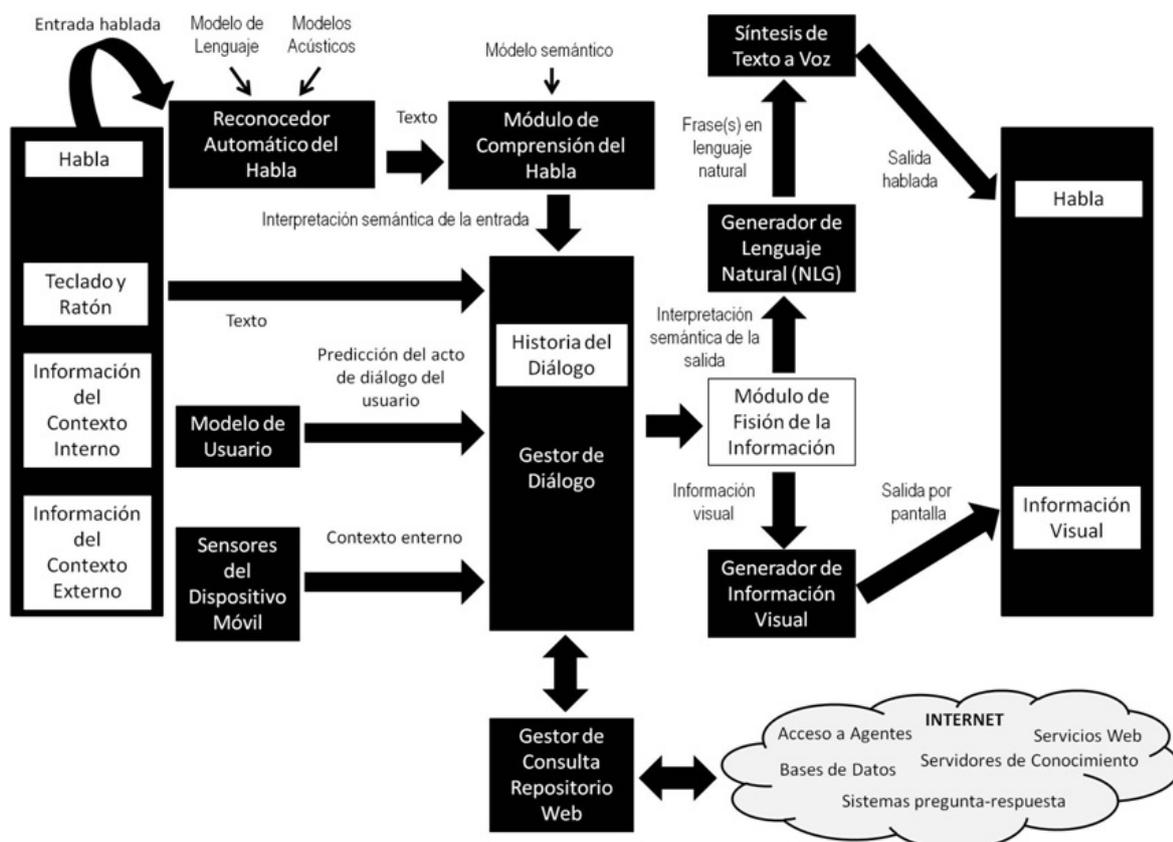
Durante el proceso de comunicación, el sistema genera un mensaje inicial, normalmente para dar la bienvenida o informar al usuario sobre las características y funcionalidades del sistema. Tras cada intervención del usuario, el sistema debe realizar un conjunto de acciones básicas que se repiten cíclicamente como respuesta a cada acción del usuario: reconocer la secuencia de palabras mencionadas por el usuario; extraer el significado de dichas palabras (es decir, comprender qué información es útil en el dominio del sistema), realizar operaciones de acceso a base de datos u otros repositorios del sistema, en los que se almacena la información que solicita el usuario o se registran las operaciones que desea conocer; decidir qué acción o acciones deben realizarse a continuación de cada solicitud del usuario, es decir, qué respuesta debe suministrar el sistema, reproducir un mensaje hablado que informe al usuario qué acción ha seleccionado el sistema.

3.1. Reconocimiento Automático del Habla

El reconocimiento de voz es el proceso de obtención del texto que corresponde a una entrada acústica. Nuestra propuesta integra la API (interfaz de programación de aplicaciones, abreviado del inglés: *Application Programming Interface*) de voz de Google para incluir la funcionalidad de reconocimiento de voz en un sistema multimodal. La funcionalidad de reconocimiento de voz está disponible en los dispositivos Android para enviar mensajes, llamar a contactos, mostrar mapas de Google Maps, visitar un sitio web, iniciar una ruta con Google Navigation, completar una búsqueda en Google, escribir una nota, enviar un correo electrónico, o escuchar música, a por mencionar algunos de sus integraciones actuales.

Además de las utilidades de reconocimiento del habla que se integran dentro de los sistemas operativos Android, existe la posibilidad de desarrollar aplicaciones Android con entrada y salida oral utilizando la API Google Speech (paquete *android.speech*). Con esta API, el reconocimiento de voz puede llevarse a cabo mediante la definición de un *Intent* (descripción abstracta que utiliza Android para requerir una operación a realizar), o mediante la creación de una instancia de la clase *SpeechRecognizer*. En ambos casos, los resultados se presentan en forma de una lista con los N-mejores resultados y sus respectivas medidas de confianza.

Figura 1. Arquitectura propuesta para la generación de sistemas de diálogo multimodales en dispositivos móviles Android



3.2. Compresión del Habla y Adquisición del Contexto Externo

Una vez que el agente conversacional ha reconocido la elocución del usuario, es necesario comprender lo que ha mencionado. El Procesamiento del Lenguaje Natural conlleva considerar conocimiento morfológico, léxico, sintáctico, semántico, pragmático y de discurso (Wu y otros, 2010). En nuestra contribución, modelamos además el contexto de la interacción como una fuente de información adicional muy valiosa que puede considerarse conjuntamente con la representación semántica de la entrada del usuario.

Proponemos la adquisición del contexto externo por medio de los sensores integrados actualmente en dispositivos móviles Android. La mayor parte de estos dispositivos incorporan sensores que miden el movimiento, la orientación y otras condiciones ambientales. Estos sensores son capaces de proporcionar datos con una alta precisión y exactitud,

y son útiles para monitorear el movimiento del dispositivo en tres dimensiones o su localización, así como cambios en el entorno cerca del dispositivo.

3.3. Adquisición del contexto interno: modelado de la intención del usuario

En cuanto al contexto interno, nuestra propuesta se basa en la visión tradicional proporcionada por la teoría de actos de diálogo, que define los actos comunicativos como intenciones u objetivos que deben cumplirse durante el diálogo. Nuestra técnica se basa en el aprendizaje automático, a partir de un corpus de diálogos, de un modelo estadístico para predecir la intención del usuario durante el diálogo. El sistema de diálogo utiliza este modelo para anticiparse a las necesidades del usuario, adaptando dinámicamente sus objetivos y proporcionándoles sugerencias, así como respondiendo inmediatamente a sus interrupciones y generando preguntas para clarificar dudas detectadas durante el diálogo.

El modelo considera tanto la historia completa del diálogo hasta el instante actual, como también la información almacenada en perfiles de usuario.

El modelo de usuario simula la intención de usuario proporcionando el siguiente acto de diálogo de usuario en el mismo formato definido para el módulo de comprensión del habla. Definimos una estructura de datos, que denominamos Registro de Usuario (UR), para almacenar la información proporcionada por el usuario durante la historia previa del diálogo. Para cada instante i , el modelo de usuario estima la intención del usuario teniendo en cuenta la secuencia de actos de diálogo que precede al instante i (UR_{i-1}), la respuesta del sistema en dicho instante (A_i) y el objetivo del diálogo (\mathcal{O}). La selección de la respuesta más probable del usuario viene dada por la ecuación:

$$\hat{U}_i = \arg \max_{U_i \in \mathcal{U}} P(U_i | UR_{i-1}, A_i, \mathcal{O})$$

donde U hace referencia al conjunto de posibles respuestas definidas para el usuario, en términos de información semántica que se proporciona al sistema.

La información contenida en UR_i es un resumen de la información proporcionada por el usuario hasta el instante i , es decir, la interpretación semántica de las elocuciones del usuario durante el diálogo y la información contenida en un perfil de usuario. Este perfil incluye información general del usuario, datos relativos a su nivel de habilidad utilizando la aplicación, así como estadísticas de uso de la misma.

3.4. Gestión de la Interacción

El gestor de diálogo decide la siguiente acción del sistema, interpretando la representación semántica de la entrada del usuario en el contexto del diálogo. La metodología que proponemos para los procesos de fusión de datos y de gestión de diálogo multimodal considera el conjunto de las fuentes de entrada de información (interacción oral, interacción visual, contexto externo, y el modelado de la intención usuario) mediante una técnica de aprendizaje automático. De igual modo que en nuestros trabajos previos sobre gestión de diálogo estadística (Griol y otros, 2014), proponemos la definición de una estructura de datos similar a la del Registro de Usuario para almacenar los valores de los actos de diálogo proporcionados por medio de las diferentes modalidades de entrada a lo largo de la historia de diálogo, que hemos denominado Registro de la Interacción (IR).

La información contenida en el IR se actualiza teniendo en cuenta los valores proporcionados por los módulos de entrada del sistema a lo largo de la historia de diálogo. Cada posición del IR puede completarse normalmente por medio de más de una modalidad de entrada. Las medidas de confianza proporcionadas por los módulos de procesamiento de cada modalidad de entrada se utilizan en caso de conflicto entre los valores proporcionados para una misma posición del registro. La selección de la siguiente respuesta del sistema se lleva a cabo mediante la ecuación:

$$\hat{A}_i = \arg \max_{A_i \in \mathcal{A}} P(A_i | IR_i, A_{i-1})$$

donde A hace referencia al conjunto de respuestas definidas para el sistema, IR_i se corresponde con el contenido del Registro de la Interacción en el instante actual (i) y A_{i-1} indica la última respuesta proporcionada por el sistema (instante $i-1$).

3.5. Generación de la salida del sistema: Fisión, Generación de Lenguaje Natural y Síntesis de Texto a Voz

El módulo de Fisión de la Información recibe como entrada una representación abstracta, independientemente de las modalidades, del objetivo seleccionado por el gestor de diálogo. Este módulo utiliza estrategias de presentación que descomponen el objetivo marcado por el gestor en tareas de presentación. También decide si una descripción de objeto debe ser presentado al usuario oralmente o gráficamente. El resultado es un script de comandos de presentación que se transmite a los módulos de Generación de Lenguaje Natural e Información Visual.

El Módulo de Generación Visual genera la disposición visual de la salida del sistema utilizando elementos de diseño gráficos creados dinámicamente. La estructura visual del interfaz de usuario (*User Interface*, UI) se define en una aplicación multimodal Android a través de *layouts*. Los *layouts* pueden definirse declarando elementos UI en XML o creando instancias de *layouts* en tiempo de ejecución.

La generación de lenguaje natural es una de las tareas más importantes del Procesamiento de Lenguaje Natural, que trata de la producción automática de textos en lenguaje natural a partir de una representación interna no-lingüística utilizada por el sistema de diálogo (Lemon, 2011). La aproximación más sencilla consiste en utilizar mensajes predefinidos a modo de plantillas de texto.

Finalmente, el sintetizador de texto a voz (*text-to-speech synthesizer*, TTS) genera la señal de voz

para transmitir este mensaje al usuario. Proponemos la utilización de la API Google TTS para incorporar la funcionalidad de síntesis de voz en una aplicación Android. El paquete *android.speech.tts* proporciona las clases e interfaces necesarios para integrar esta funcionalidad. Todo dispositivo Android incorpora un motor TTS. Además, Android permite la instalación y personalización de varios motores.

4. CALIMACO: SISTEMA DE DIÁLOGO MULTIMODAL PARA LA PROVISIÓN DE SERVICIOS BIBLIOTECARIOS ADAPTADOS MEDIANTE DISPOSITIVOS MÓVILES

La aplicación Android desarrollada consiste en un asistente virtual multimodal cuya función es la de proporcionar, a través de un sistema de diálogo multimodal, información relativa a un servicio de biblioteca digital atendiendo a los gustos y preferencias del usuario. La aplicación recibe el nombre de Calimaco en honor a Calimaco de Cirene, director de la Biblioteca de Alejandría hasta el 240 A.C. y creador de los *pinakes* en el 260 A.C. (primeros catálogos de libros que sirvieron como índices de contenidos de la biblioteca).

Al ser una aplicación multimodal, la aplicación permite al usuario introducir su petición mediante el modo táctil, mediante la voz o combinando ambos modos. Las siguientes subsecciones describen las tecnologías utilizadas para desarrollar la aplicación y principales funcionalidades que ofrece.

4.1. Tecnologías utilizadas

La aplicación se ha desarrollado de acuerdo a la arquitectura descrita en la Figura 1. En cuanto al reconocimiento automático del habla, tal y como se ha comentado anteriormente, la mayor parte de asistentes de voz existentes utilizan el servicio de voz ofrecido por Google, ya que ofrece altas prestaciones además de tener una fácil integración en las aplicaciones Android.

A la hora de elegir el motor de síntesis de voz integrado en la aplicación se ha tenido en cuenta la calidad de las voces que ofrece, los idiomas que tiene disponibles para Android, el tamaño de la descarga, la disponibilidad y el precio. Tras el estudio de los motores existentes, se ha decidido emplear el motor de síntesis IVONA TTS HQ, ya que ofrece multitud de voces femeninas y masculinas de gran calidad y es actualmente gratuito.

Los modelos estadísticos para los módulos del reconocimiento de la intención del usuario y de gestión de diálogo se aprendieron a partir de un corpus adquirido por medio de una técnica de generación automática de diálogo desarrollada pre-

viamente (Griol y otros, 2013). La aplicación permite además que el usuario registre su nombre de usuario y contraseña. Además, puede completar un formulario con sus gustos y preferencias. Posteriormente, la aplicación Android (cliente) envía dicha información a un servidor web que accede a una base de datos MySQL remota con el objetivo de almacenar los datos de usuario en una tabla de usuarios registrados.

La aplicación utiliza diferentes bases de datos adicionales a modo de repositorio de la información, varias de ellas dispuestas en un servidor web externo para posibilitar así el acceso a la aplicación y la identificación del usuario desde más de un terminal móvil.

Actualmente, la API de Android no provee ningún mecanismo que permita acceder directamente a través de Internet a una base de datos MySQL remota y ejecutar una consulta dentro de ella. Para ello, ha sido necesario desarrollar un servicio web específico. Mediante este servicio, la aplicación Android envía una consulta mediante HTTP a un servidor web que devuelve la información solicitada en formato JSON (JavaScript Object Notation).

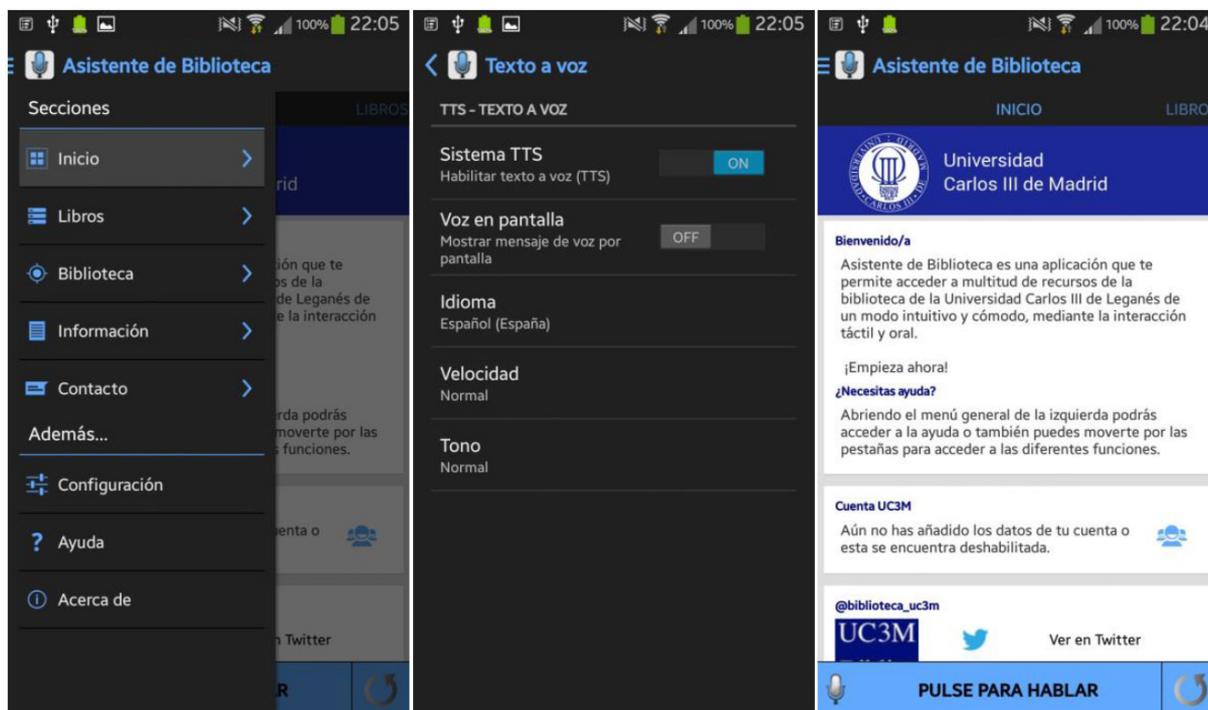
La aplicación incluye además funcionalidades que requieren la consulta de información contenida en páginas web (por ejemplo, para acceder al catálogo de libros o consultar la regulación de la biblioteca). Para realizar esta consulta, la aplicación Android se conecta mediante HTTP al servidor web que contiene dicha información y la extrae del contenido HTML mediante la utilización de la librería de Java JSoup. Finalmente, la aplicación desarrollada utiliza además las API Google Maps, Google Directions y Google Places.

4.2. Funcionalidades del sistema

Mediante la arquitectura y tecnologías descritas en las secciones anteriores, la aplicación desarrollada proporciona cuatro bloques fundamentales de funcionalidades:

1. *Registro en la aplicación e Información general:* Este primer bloque posibilita registrarse en la aplicación (necesario para realizar la adaptación al usuario y recopilar las consultas más frecuentes realizadas por cada uno de ellos) y configurar una amplia gama de opciones relacionadas con sus funcionalidades (ver Figura 2). Además, es posible acceder a los recursos de la Web 2.0 utilizados por la Universidad Carlos III de Madrid (página web de la Biblioteca de la UC3M, cuenta Twitter, página de Facebook y repositorio institucional de noticias, agenda semanal y novedades).

Figura 2. Capturas de pantalla de la funcionalidad de Configuración de la aplicación e Información general



2. *Acceso al Catálogo de la Biblioteca:* La aplicación facilita el acceso al catálogo de la biblioteca almacenado en las bases de datos MySQL correspondientes. Los usuarios pueden proporcionar el título, autor o materia que desea consultar. Como respuesta, la aplicación proporciona la ficha completa del recurso solicitado, la información correspondiente a su estado (en préstamo, prestado o en lista de espera) e imágenes disponibles en el repositorio web de la biblioteca de la universidad (ver Figura 3).
3. *Guiado por la Biblioteca:* La aplicación además guía al usuario por la Biblioteca de la Universidad Carlos III de Madrid en el Campus de Leganés. Para ello, se debe proporcionar la ubicación a la que desea ser guiado (sala concreta o servicio, materia del catálogo o título del recurso). La aplicación usa planos, fotografías y texto sintetizado que transmiten las indicaciones para localizar la ubicación requerida. La Figura 4 muestra diferentes capturas de pantalla de esta funcionalidad de la aplicación.
4. *Preguntas frecuentes:* Para el diseño de la aplicación se consultó las preguntas más frecuentes realizadas por los usuarios de las bibliotecas universitarias de la Universidad de León¹, Universidad de Alcalá², Universidad de Granada³, Universidad Carlos III de Madrid⁴, Universidad de Almería⁵, Universidad de Murcia⁶, Universidad Autónoma de Barcelona⁷, Universidad de Córdoba⁸ y la Universidad de Alicante⁹. El conjunto de preguntas se han clasificado en 13 categorías (General, Carné Universitario, Servicios e Infraestructuras, Catálogo, Préstamo, Búsqueda de Información y Acceso a los Recursos, Buscador, Revistas electrónicas, Libros electrónicos, Repositorio Institucional, Cursos de Formación, Impacto y Citas, Adquisiciones), que engloban un total de más de 250 consultas frecuentes distintas efectuadas por los usuarios de dichas bibliotecas.
5. *Contacto con la biblioteca:* Esta funcionalidad proporciona el acceso a un formulario web en el que los usuarios pueden solicitar información directamente al personal de la universidad.

Figura 3. Capturas de pantalla de la funcionalidad de Acceso al Catálogo



Figura 4. Capturas de pantalla de la funcionalidad de Guiado por la Biblioteca

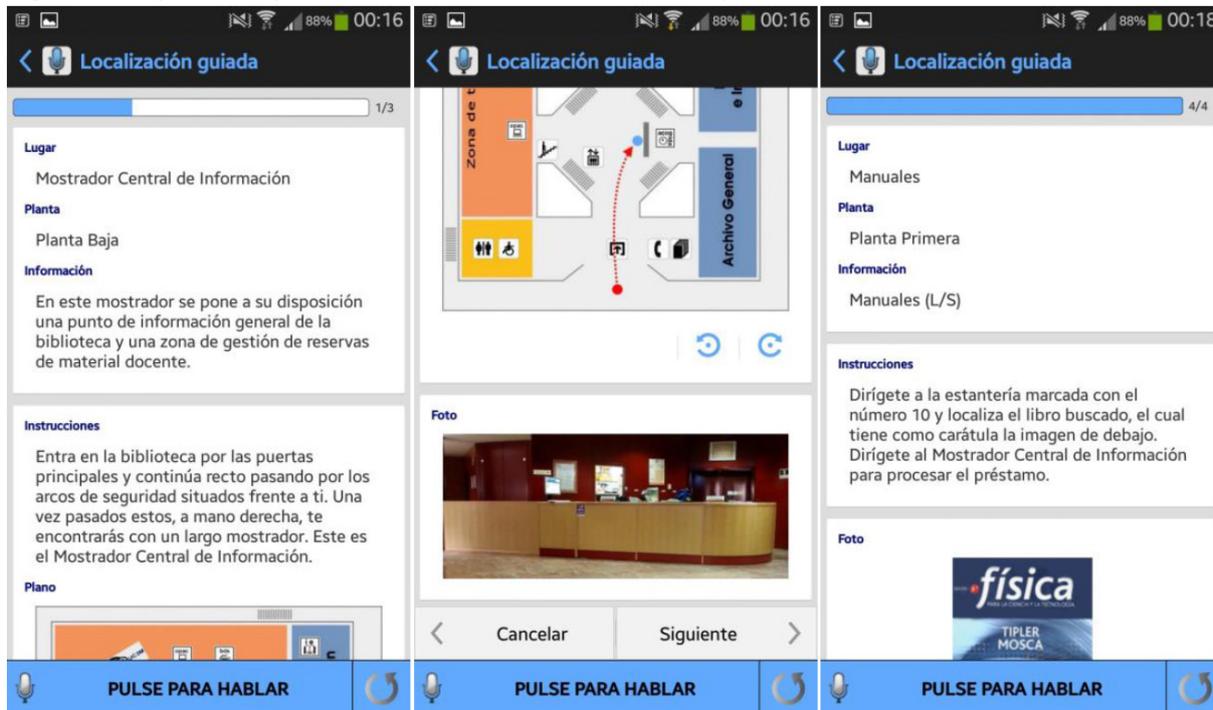
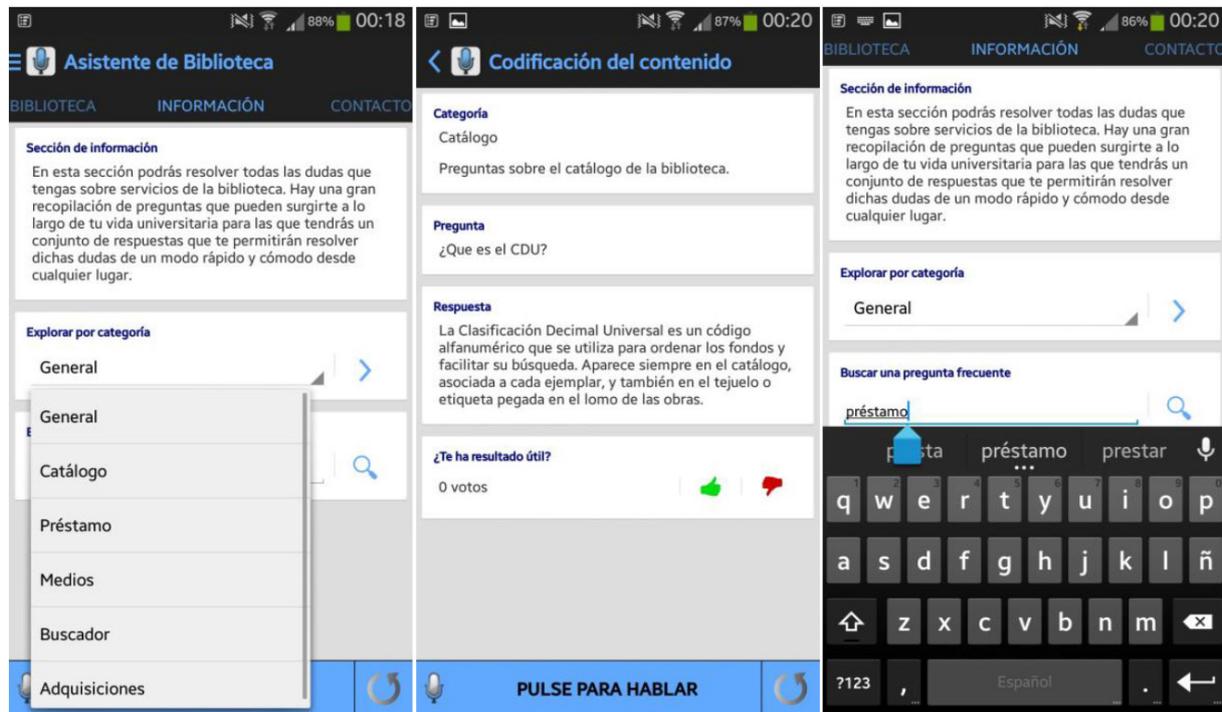


Figura 5. Ejemplo de consulta frecuente y respuesta proporcionada por el sistema

¿Cómo funciona la base de datos de revistas?
En las fichas de los recursos electrónicos encontrará las guías de uso y tutoriales que les ofrecemos, en muchos casos elaborados por la misma biblioteca. También puede inscribirse a los cursos que periódicamente programan las bibliotecas.

Figura 6. Capturas de pantalla de la funcionalidad de Preguntas Frecuentes

sobre consultas que no hayan podido resolver a través del conjunto de funcionalidades descritas anteriormente, además de servir para la transmisión de sugerencias, quejas o felicitaciones. La Figura 7 muestra tres capturas de pantalla correspondientes a esta funcionalidad de la aplicación.

6. *Ayuda y Búsqueda Rápida:* Este último bloque de la aplicación muestra una ayuda detallada sobre las diferentes funcionalidades de la aplicación. Esta información se proporciona multimodalmente mediante capturas de pantalla de la aplicación, vídeos didácticos con ejemplos prácticos de uso de cada una de las funcionalidades descritas y textos sintetizados a voz. La Figura 8 muestra capturas de pantalla de esta funcionalidad de la aplicación.

5. EVALUACIÓN

La metodología utilizada para evaluar la aplicación multimodal se basa en un conjunto de estudios relativos a la evaluación fiable de sistemas de diálogo multimodales (Metze y otros, 2009). Según dichos estudios, todo desarrollo y evaluación de un sistema de diálogo multimodal requiere valorar el rendimiento del sistema en sus distintas modalidades. Para ello, se propone un análisis de

la usabilidad del sistema a través de cuestionarios de opinión, cuyas preguntas están orientadas a evaluar el sistema multimodal además de la interacción mediante el modo táctil o mediante la voz. La Tabla I muestra el cuestionario que hemos elaborado para evaluar la aplicación en sus diferentes modalidades.

El cuestionario consta de 9 preguntas, en las que las preguntas 3-9 se han evaluado para usuarios interactuando con la aplicación mediante el modo táctil, el modo oral y el modo multimodal combinando ambas modalidades de entrada y de salida. Cada pregunta tiene 5 posibles respuestas de las que sólo se puede elegir una de ellas. Tal y como se observa, los aspectos que se han querido analizar para las distintas modalidades del sistema son: el grado en el cual el usuario valora que es entendido por el sistema y entiende los mensajes del mismo, la velocidad percibida de la de interacción, el nivel de dificultad del sistema, la presencia de errores, la seguridad de lo que se debe hacer en cada momento, y el nivel de satisfacción con el sistema global. Además, la información adicional de los usuarios sobre su grado de conocimiento y experiencia previa con los interfaces multimodales y los interfaces orales sirve para hacerse una idea del perfil de los usuarios.

Figura 7. Capturas de pantalla de la funcionalidad Contacto con la biblioteca

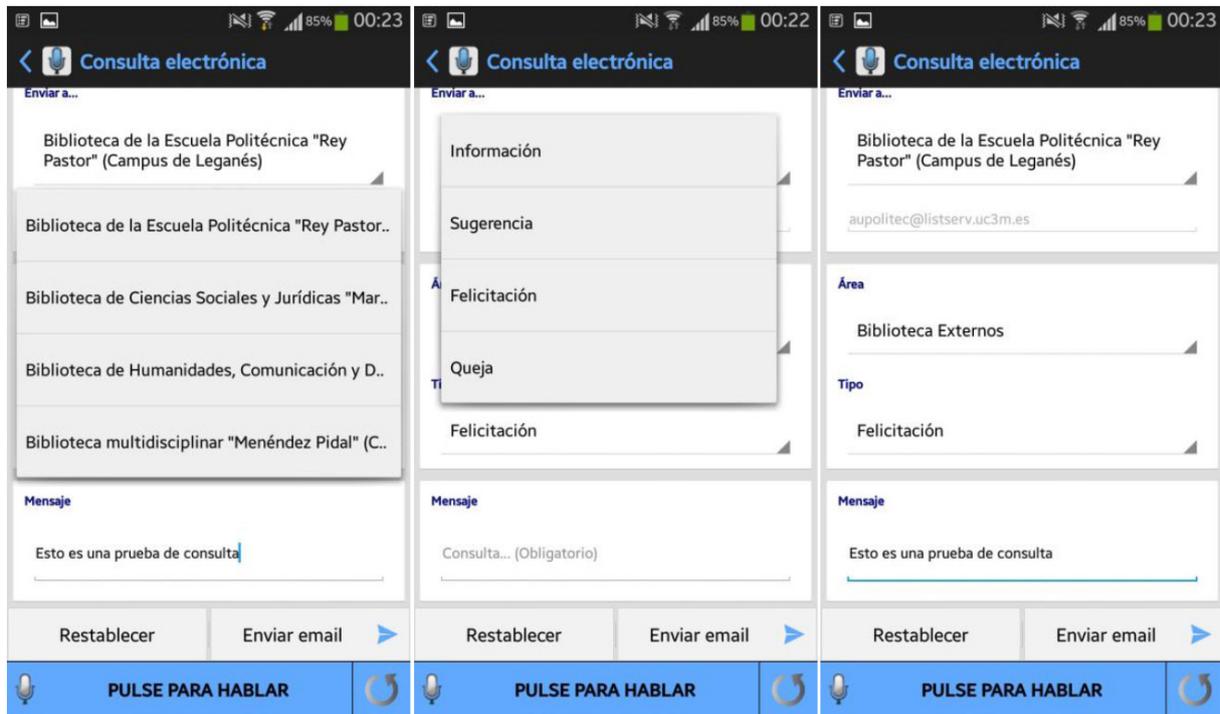


Figura 8. Capturas de pantalla de la funcionalidad Ayuda y Búsqueda Rápida

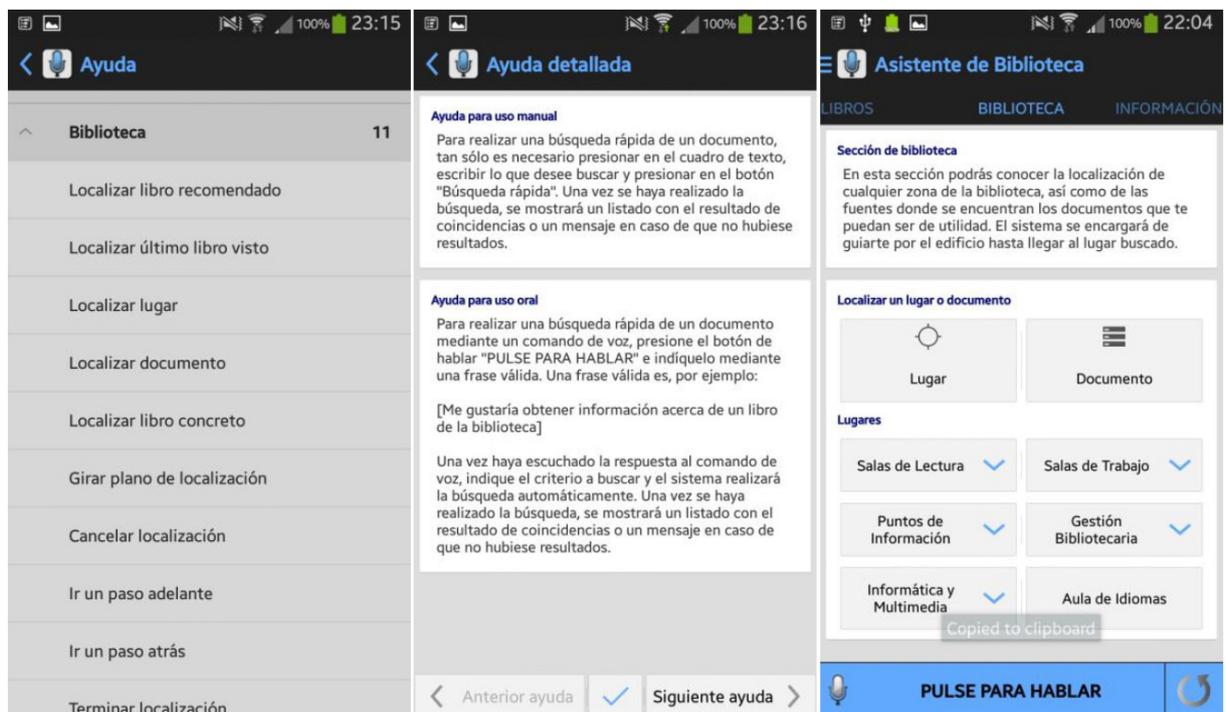


Tabla I. Cuestionario de evaluación utilizado para la evaluación de la aplicación

Preguntas	Posibles respuestas
1. Puntúe en una escala de 1 a 5 su experiencia previa usando Interfaces de voz	1, 2, 3, 4, 5
2. Puntúe en una escala de 1 a 5 su experiencia previa usando Interfaces multimodales	1, 2, 3, 4, 5
3. ¿Qué tal ha entendido el sistema sus peticiones?	Muy mal, Mal, Regular, Bien, Muy bien
4. ¿Cómo ha entendido los mensajes generados por el sistema?	Muy mal, Mal, Regular, Bien, Muy bien
5. En su opinión la interacción fue...	Muy lenta, Lenta, Adecuada, Rápida, Muy rápida
6. Establezca el nivel de dificultad del sistema	Muy difícil, Difícil, Normal, Fácil, Muy fácil
7. ¿Ha percibido errores durante su interacción con el sistema?	Sí, he percibido muchos errores lo que ha imposibilitado la interacción con el sistema Sí, he percibido bastantes errores lo que ha causado gran dificultad en la interacción con el sistema Sí, he percibido algunos errores lo que ha causado cierta dificultad en la interacción con el sistema Sí, he percibido algunos errores aunque la interacción con el sistema no se ha visto afectada en absoluto No he percibido errores
8. ¿Le ha sido fácil decidir qué hacer en cada momento para obtener la información que usted solicitaba?	No, ha sido imposible No, me ha supuesto gran dificultad Sí, pero con ciertas dificultades Sí, ha sido fácil Sí, ha sido muy fácil
9. En términos generales, ¿está usted satisfecho con el sistema?	No, nada Poco satisfecho Satisfecho Bastante satisfecho Muy satisfecho

En la evaluación han participado 25 usuarios hispanoparlantes (12 hombres, 13 mujeres) con edad comprendida entre 22 y 54 años. En primer lugar, los usuarios evaluaron el sistema utilizando el modo táctil. Posteriormente, evaluaron el sistema utilizando los interfaces orales. Finalmente, evaluaron el sistema multimodal combinando la utilización de los interfaces tradicionales (pantalla y teclado) con los interfaces orales. Los usuarios eligieron libremente las acciones a realizar durante la evaluación de la aplicación por lo que los módulos y submódulos a los que accedieron fueron variados.

La Tabla II recopila las estadísticas correspondientes a las distintas preguntas y aspectos evaluados. En cuanto a los resultados obtenidos en las preguntas referidas a la evaluación del nivel de experiencia previa de los usuarios con interfaces orales y multimodales, se concluye que el nivel de conocimiento de los usuarios a los que se les ha realizado la encuesta es variado, siendo mayor el

número de usuarios que poseen un conocimiento medio en la utilización de ambos modos de acceso. Se ha intentado seleccionar tanto a usuarios familiarizados con los interfaces multimodales y orales como a usuarios que no tenían experiencia previa con el objetivo de extraer conclusiones más fiables.

En cuanto al grado en el cual el usuario valora que es entendido por el sistema, para el modo táctil se esperaba obtener estos resultados de antemano. Para el modo oral y el modo multimodal estos resultados han sido posibles gracias a la eficacia del sistema de reconocimiento de voz de Google, que está en continua evolución y es capaz en la actualidad, de reconocer perfectamente cualquier secuencia de voz que el usuario emita. Además, para mejorar esta característica en el módulo búsqueda de material bibliográfico, se ha tenido en cuenta que el título pueda no estar traducido, por lo que se permite al usuario pronunciar el título en su idioma original, seleccionando por adelantado dicho idioma.

Tabla II. Resultados de la evaluación

Aspecto evaluado		Resultados obtenidos			
		Valor máximo	Valor mínimo	Desviación típica	
Valor medio					
Experiencia previa usando Interfaces de voz		3.36	5	1	1.26
Experiencia previa usando Interfaces multimodales		3.24	5	1	1.34
Grado en el cual el usuario valora que es entendido por el sistema	Modo táctil	4.80	5	4	0.4
	Modo oral	4.60	5	4	0.48
	Modo multimodal	4.84	5	4	0.37
Grado en el cual el usuario valora que entiende los mensajes generados por el sistema	Modo táctil	4.92	5	4	0.27
	Modo oral	3.90	5	3	0.69
	Modo multimodal	4.48	5	4	0.5
Velocidad percibida de la interacción.	Modo táctil	3.64	5	3	0.69
	Modo oral	2.8	5	2	1.02
	Modo multimodal	4.76	5	3	0.51
Nivel de dificultad del sistema.	Modo táctil	4.24	5	3	0.51
	Modo oral	3.6	5	3	0.69
	Modo multimodal	4.84	5	4	0.37
Presencia de errores	Modo táctil	4.8	5	4	0.4
	Modo oral	3.8	5	2	0.98
	Modo multimodal	4.44	5	4	0.5
Seguridad del usuario en lo que debe hacer en cada momento	Modo táctil	4	5	3	0.49
	Modo oral	4.4	5	3	0.63
	Modo multimodal	4.84	5	4	0.37
Nivel de satisfacción global con el sistema	Modo táctil	4.44	5	4	0.5
	Modo oral	3.64	5	3	0.74
	Modo multimodal	4.88	5	4	0.32

Con respecto al grado en el cual el usuario valora que entiende los mensajes generados por el sistema, antes de la evaluación de esta característica, se esperaba que el modo multimodal obtuviese los mejores resultados de la encuesta ya que permite combinar el modo táctil con el oral y por lo tanto aprovechar las ventajas de ambos. Sin embargo, a pesar de que los usuarios hayan valorado en general que entienden muy bien los mensajes generados por el sistema en el modo multimodal, el modo que ha obtenido los mejores resultados ha sido el modo táctil, seguido del modo multimodal y finalmente del modo oral, siendo este último el que ha obtenido los peores resultados.

En cuanto al modo táctil no se esperaba obtener tan buenos resultados, ya que se suponía de antemano que los usuarios con problemas de visión iban a tener mayores dificultades para leer la información en la pantalla del dispositivo y que les fuese más cómodo por lo tanto utilizar el modo oral o el modo multimodal. Sin embargo, al diseñar la

aplicación se intentó que los resultados se mostrasen de forma clara por pantalla, utilizando un tamaño de fuente grande, por lo que todos los usuarios han entendido a la perfección los mensajes que ha generado el sistema. El modo oral ha obtenido resultados satisfactorios aunque peores que los esperados. Los resultados en el modo oral se han visto afectados negativamente por dos factores:

- El grado en el cual el usuario valora que entiende los mensajes generados por el sistema depende de la calidad de las voces que el usuario tenga instaladas en el dispositivo. Tal y como se ha visto en el estudio efectuado al principio de este proyecto, el sintetizador de voz IVONA TTS HQ es el que ofrece las voces de mayor calidad. Sin embargo, algunos usuarios no han optado por descargarse sus voces y han mantenido la voz que les viene por defecto en su dispositivo o la que pide instalar Google por defecto si no encuentra ninguna voz en la aplicación.

- Cuando el sintetizador está leyendo un texto en un determinado idioma, no diferencia las cadenas dentro del texto que están en otro idioma por lo que las pronuncia en el idioma en el que se estaba efectuando la síntesis de texto a voz. Esto se podría solucionar ya que el diseñador puede especificar la pronunciación de determinadas cadenas de texto. Sin embargo, es muy difícil tener en cuenta todas las posibles cadenas existentes y el texto podría escucharse entrecortado. El módulo recomendación de material bibliográfico también se ve afectado por el problema de pronunciación en un idioma incorrecto cuando se encuentran títulos no traducidos. No obstante, en el módulo búsqueda de material bibliográfico, el sintetizador de voz lee la información del recurso en el idioma especificado por el usuario, evitando este problema.

En cuanto a la velocidad percibida de la interacción, los usuarios han valorado que la velocidad de la interacción en el modo multimodal ha sido muy rápida, en el modo táctil ha sido buena y en el modo oral ha sido adecuada. Cabe destacar que la velocidad en interacción oral entre el usuario y la aplicación se ha visto dificultada al no existir un mecanismo en Android para que la síntesis de voz se detenga cuando el usuario desee comenzar a hablar. Esto resulta incómodo para el usuario considerando que este va aprendiendo cómo funciona la aplicación según va utilizándola, y no le resulta necesario esperar a que termine de hablar la aplicación para emitirle sus peticiones a través del reconocedor de voz. El modo multimodal ha obtenido los mejores resultados para esta característica ya que el usuario puede seleccionar la modalidad que le resulte más cómoda y rápida de utilizar en todo momento. Además, puede pulsar un botón para pausar la síntesis y un botón para iniciar el reconocimiento de voz, evitando por lo tanto la desventaja principal del modo oral.

Con respecto al nivel de dificultad del sistema, los usuarios han valorado que la utilización de la aplicación en modo multimodal ha sido la más fácil de utilizar. También han valorado la interacción con la aplicación en modo táctil como muy fácil y la interacción en modo oral como fácil. El modo multimodal ha obtenido los mejores resultados para esta característica ya que el usuario puede seleccionar la modalidad que le resulte más cómoda y fácil de utilizar en todo momento, sin estar restringido a las dificultades que pueda encontrar en un único modo y evitar de esta forma quedarse atascado sin saber qué decisión to-

mar. En el modo multimodal y oral se emiten por voz instrucciones de utilización de la aplicación lo que facilita la interacción con la misma. El modo táctil no muestra por pantalla las instrucciones de utilización por lo que el usuario ha de apoyarse únicamente en su intuición a la hora de decidir qué decisión tomar.

En cuanto a la presencia de errores, los usuarios no han percibido en general errores en su interacción con la aplicación en modo táctil. En el modo multimodal y oral han percibido algunos errores pero que no han dificultado en absoluto su interacción con la aplicación. Aún así, el modo oral es el que se ha visto más afectado ante la presencia de errores ya que el usuario puede sentirse más inseguro a la hora de tomar una decisión después de la ocurrencia de un error al no haber aun un mecanismo de retroalimentación lo suficientemente logrado en el modo oral. Por lo general, el usuario valora que es más fácil apoyarse en lo que ve por pantalla cuando ha sucedido algún error.

Con respecto a la seguridad del usuario en lo que debe hacer en cada momento, los usuarios han sentido una mayor seguridad en lo que debían hacer en cada momento con la interacción en modo multimodal, seguido de la interacción en modo oral y finalmente del modo táctil. A pesar de ello, ha sido fácil tomar decisiones en los tres modos de acceso a la aplicación. El motivo principal por el cual el modo multimodal y el modo oral resultan más ventajosos en comparativa con el modo táctil en cuanto a esta característica es que, tal y como se ha explicado anteriormente, la aplicación en estos modos emite instrucciones de utilización de la aplicación y cuenta con más mecanismos de ayuda. El modo táctil se apoya únicamente en la intuición que tiene el usuario a la hora de decidir qué acción tomar. Aun así, el usuario se siente más cómodo utilizando el modo multimodal que el modo oral ya que si no entiende alguna instrucción u ocurre algún error puede apoyarse también en lo que ve por pantalla.

Finalmente, con respecto al nivel de satisfacción con el sistema global, los usuarios de la aplicación encuestados se han sentido muy satisfechos con el modo multimodal y táctil, y satisfechos con el modo oral. El modo multimodal ha sido el mejor valorado por los usuarios ya que como se desprende de las conclusiones extraídas del resto de preguntas de la encuesta, el usuario prefiere combinar el modo táctil y el modo oral según le resulte más eficaz, aprovechando todas las ventajas y evitando todas las desventajas de ambos modos.

6. CONCLUSIONES Y TRABAJO FUTURO

Los sistemas interactivos multimodales ofrecen al usuario una combinación de modalidades de entrada y salida para interactuar con sus dispositivos, pudiendo aprovechar las ventajas que proporcionan modalidades más naturales de acceso a la información como es el habla. Diferentes proveedores ofrecen APIs para el desarrollo de aplicaciones que utilizan el habla como una posible modalidad de entrada y de salida, pero los desarrolladores tienen que diseñar soluciones ad-hoc para poner en práctica la gestión de la interacción. En nuestro trabajo aportamos una arquitectura para desarrollar agentes conversacionales multimodales sensibles al contexto que se pueden integrar fácilmente en dispositivos móviles Android. La arquitectura engloba un conjunto de módulos que cooperan para proporcionar servicios adaptados y definen un modo de representación para el intercambio de conocimiento entre los diferentes componentes de la arquitectura.

En nuestra propuesta, la interacción se gestiona dinámicamente mediante un proceso de clasificación que selecciona la mejor respuesta para el agente, teniendo en cuenta la historia previa de la interacción. Hemos adaptado esta metodología para desarrollar un gestor de diálogo sensible al contexto que utiliza una estructura de datos que almacena no sólo la información proporcionada por el usuario con respecto a la tarea específica del sistema, sino además la información proporcionada por un reconocedor de la intención del usuario y el contexto externo proporcionado por los sensores en el dispositivo móvil. Para almacenar la información de contexto y la intención hemos definido una estructura de datos que también permite gestionar perfiles de usuario. Estos perfiles incluyen información acerca de sus preferencias y necesidades, algunas de ellas extraídas automáticamente de las interacciones previas con la aplicación. Mediante la arquitectura propuesta es posible desarrollar interfaces multimodales que optimizan la gestión de la interacción e integran diferentes fuentes de información que posibilitan que la aplicación se adapte al usuario y al contexto de la interacción.

Para mostrar el potencial de nuestra propuesta, hemos implementado una aplicación multimodal Android que proporciona servicios bibliotecarios

avanzados, que incorporan dinámicamente necesidades específicas del usuario y sus preferencias, así como las características específicas del entorno en el que se produce interacción, con el fin de mejorar y personalizar el servicio que se proporciona mediante un dispositivo móvil con acceso a Internet. Para desarrollar este sistema, hemos definido los requisitos completos para la tarea y desarrollado los diferentes módulos, y la información necesaria incorporada en los perfiles de usuario.

Hemos completado una evaluación del agente desarrollado para evaluar los beneficios de la interacción multimodal y la sensibilidad al contexto. En cuanto a la interacción multimodal, los usuarios han evaluado la aplicación utilizando únicamente el interfaz táctil-visual, sólo la voz y el modo multimodal. Los resultados muestran que las tasas máximas de satisfacción se alcanzan con el modo multimodal.

Como trabajo futuro, estamos ampliando actualmente las funcionalidades de adaptación de la aplicación al usuario, utilizando un registro para almacenar las interacciones previas con la misma y utilizando esta información para adaptar el funcionamiento de la aplicación según las preferencias específicas de cada usuario, errores detectados en la interacción y consultas más frecuentes. Con los resultados de estas actividades, optimizaremos el sistema para posibilitar su disponibilidad en Google Play, así como en el repositorio de herramientas multimedia de nuestra universidad. Queremos además ampliar la aplicación para posibilitar su uso en dispositivos móviles con otros sistemas operativos, especialmente iOS y Windows Phone.

7. AGRADECIMIENTOS

Trabajo parcialmente financiado por los proyectos MINECO TEC2012-37832-C02-01, CICYT TEC2011-28626-C02-02, CAM CONTEXTS (S2009/TIC-1485).

ACKNOWLEDGEMENTS

Work partially funded by Projects MINECO TEC2012-37832-C02-01, CICYT TEC2011-28626-C02-02, CAM CONTEXTS (S2009/TIC-1485).

8. NOTAS

- [1] <http://biblioteca.unileon.es/>
 [2] <http://www.uah.es/biblioteca/>
 [3] <http://biblioteca.ugr.es/>
 [4] <http://biblioteca.uc3m.es/>
 [5] <http://www.ual.es/biblioteca>
 [6] <http://www.um.es/web/biblioteca/contenido/informacion/preguntas-mas-frecuentes>

- [7] <http://www.uab.cat/web/recursos-de-informacion/preguntas-mas-frecuentes-y-guias-1260947893092.html>
 [8] <http://www.uco.es/servicios/biblioteca/buco/faq.index.html>
 [9] <http://biblioteca.ua.es/es/faqs/preguntas-mas-frecuentes.html>

9. REFERENCIAS

- Balagué, N.; Dören, P.; Juntunen, A.; Saarti, J. (2014). Quality Audits as a Tool for Quality Improvement in Selected European Higher Education Libraries. *The Journal of Academic Librarianship*, vol. 40 (5), 529-533. <http://dx.doi.org/10.1016/j.acalib.2014.01.002>
- Becker, R.; Cáceres, R.; Hanson, K.; Isaacman, S., Loh, J., Martonosi, M., Rowland, J., Urbanek, S.; Varshavsky, A.; Volinsky, C. (2013). Human mobility characterization from cellular network data. *Communications of the ACM*, vol. 56(1), 74-82. <http://dx.doi.org/10.1145/2398356.2398375>
- Calhoun, K. (2002). From information gateway to digital library management system: a case analysis. *Library Collections, Acquisitions, and Technical Services*, vol. 26(2), 141-150. [http://dx.doi.org/10.1016/S1464-9055\(02\)00226-9](http://dx.doi.org/10.1016/S1464-9055(02)00226-9)
- Calhoun, K.; Cellentani, D. (2009). Online catalogers: What users and librarians want. *OCLC report*. Dublin, OH: OCLC.
- Casey, M.E.; Savastinuk, L.C. (2007). Library 2.0. A Guide to Participatory Library Service, *Information Today*.
- Castaño, B.; E-Martín, Y.; R-Moreno, M.D.; Usero, L. (2013). Sistema Inteligente de Detección y Orientación de usuarios en Bibliotecas. *Revista Española de Documentación Científica*, 36(1):en003. <http://dx.doi.org/10.3989/redc.2013.1.916>
- Chen, H.L.; Albee, B. (2012). An open source library system and public library users: Finding and using library collections. *Library & Information Science Research*, vol. 34(3), 220-227. <http://dx.doi.org/10.1016/j.lisr.2011.12.001>
- Chen, C.M.; Tsai, Y.N. (2012). Interactive augmented reality system for enhancing library instruction in elementary schools. *Computers & Education*, vol. 59(2), 638-652. <http://dx.doi.org/10.1016/j.compedu.2012.03.001>
- Díaz-Valenzuela, I.; Martín-Bautista, M.J.; Vila, M.A.; Campaña, J.R. (2013). An automatic system for identifying authorities in digital libraries. *Expert Systems with Applications*, vol. 40 (10), 3994-4002. <http://dx.doi.org/10.1016/j.eswa.2013.01.010>
- Dwivedi, Y.K.; Kapoor, K.K.; Williams, M.D.; Williams, J. (2013). RFID systems in libraries: An empirical examination of factors affecting system use and user satisfaction. *International Journal of Information Management*, vol. 33(2), 367-377. <http://dx.doi.org/10.1016/j.ijinfomgt.2012.10.008>
- Evergreen. (2010). Evergreen Indiana: Resource & sharing cooperative of Evergreen libraries. Disponible en: <http://rscel.evergreen-ils.org/node/1457>
- Fagan, J.C. (2014). The Suitability of Web Analytics Key Performance Indicators in the Academic Library Environment. *The Journal of Academic Librarianship*, vol. 40(1), 25-34. <http://dx.doi.org/10.1016/j.acalib.2013.06.005>
- González-Fernández-Villavicencio, N.; Cánovas-Álvarez, E.; Arahal-Junco, C. (2014). Evaluación del servicio de referencia de una biblioteca universitaria: Biblioteca de la Universidad de Sevilla. *Revista Española de Documentación Científica*, 37(2):e045. <http://dx.doi.org/10.3989/redc.2014.2.1072>
- Griol, D.; Callejas, Z.; López-Cózar, R.; Riccardi, G. (2014). A domain-independent statistical methodology for dialog management in spoken dialog systems. *Computer Speech and Language*, vol. 28(3), 743-768. <http://dx.doi.org/10.1016/j.csl.2013.09.002>
- Griol, D.; Carbó, J.; Molina, J. (2013). An Automatic Dialog Simulation Technique to Develop and Evaluate Interactive Conversational Agents. *Applied Artificial Intelligence*, vol. 27(9), 759-780. <http://dx.doi.org/10.1080/08839514.2013.835230>
- Helling, J. (2010). A case study of one library's decision to migrate to an open source ILS. *Library Review*, vol. 59(9), 707-707.
- Kaufmann, T.; Pfister, B. (2012). Syntactic language modeling with formal grammars. *Speech Communication*, vol. 54(6), 715-731.
- Larson, J.A. (2002). VoiceXML Introduction to Developing Speech Applications. Prentice Hall.
- Lemon, O. (2011). Learning what to say and how to say it: Joint optimisation of spoken dialogue

- management and natural language generation. *Computer Speech & Language*, vol. 25(2), 210–221. <http://dx.doi.org/10.1016/j.csl.2010.04.005>
- Liu, Z. (2006) Print vs. electronic resources: a study of user perceptions, preferences, and use. *Information Processing and Management*, vol. 42(2), 583–592. <http://dx.doi.org/10.1016/j.ipm.2004.12.002>
- Mann, T. (2008). On the record but off the track: A review of the report of the Library of Congress Working Group on Bibliographic Control, with a further examination of Library of Congress cataloging tendencies. Disponible en: <http://www.guild2910.org/WorkingGrpResponse2008.pdf>
- McTear, M., Callejas, Z. (2013). Voice Application Development for Android. Packt Publishing.
- Metze, F.; Wechsung, I.; Schaffer, S.; Seebode, J.; Möller, S. (2009). Human-Computer Interaction. Novel Interaction Methods and Techniques, Springer-Verlag, cap. Reliable Evaluation of Multimodal Dialogue Systems, pp 75–83. http://dx.doi.org/10.1007/978-3-642-02577-8_9
- Moreno, M.D.; Castaño, B.; Barrero, D.F.; Hellín, A.M. (2014). Efficient Services Management in Libraries using AI and Wireless techniques. *Expert Systems with Applications*, vol. 41(17), 7904–7913. <http://dx.doi.org/10.1016/j.eswa.2014.06.047>
- O'Reilly, T. (2005). What is Web 2.0. Design Patterns and Business Models for the Next Generation of Software.
- Orduña-Malea, E. (2013). Espacio universitario español en la Web (2010): estudio descriptivo de instituciones y productos académicos a través del análisis de subdominios y subdirectorios. *Revista Española de Documentación Científica*, 36(3):e017. <http://dx.doi.org/10.3989/redc.2013.3.958>
- Paek, T.; Pieraccini, R. (2008). Automating spoken dialogue management design using machine learning: An industry perspective. *Speech Communication*, vol. 50(8–9), 716–729. <http://dx.doi.org/10.1016/j.specom.2008.03.010>
- Park, N.; Roman, R.; Lee, S.; Chung, J.E. (2009). User acceptance of a digital library system in developing countries: An application of the Technology Acceptance Model. *International Journal of Information Management*, vol. 29(3), 196–209. <http://dx.doi.org/10.1016/j.ijinfomgt.2008.07.001>
- Pérez-Marín, D.; Pascual-Nieto, I. (2011). Conversational Agents and Natural Language Interaction: Techniques and Effective Practices. IGI Global. <http://dx.doi.org/10.4018/978-1-60960-617-6>
- Pieraccini, R. (2012). The Voice in the Machine: Building Computers that Understand Speech. MIT Press.
- Porcel, C.; Herrera-Viedma, E. (2010). Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries. *Knowledge-Based Systems*, vol. 23(1), 32–39. <http://dx.doi.org/10.1016/j.knsys.2009.07.007>
- Porcel, C.; Moreno, J.M.; Herrera-Viedma, E. (2009). A multi-disciplinary recommender system to advice research resources in University Digital Libraries. *Expert Systems with Applications*, vol. 36(10), 12520–12528. <http://dx.doi.org/10.1016/j.eswa.2009.04.038>
- Renda, M.E.; Straccia, U. (2005). A personalized collaborative digital library environment: a model and an application, *Information Processing and Management*, vol. 41(1), 5–21. <http://dx.doi.org/10.1016/j.ipm.2004.04.007>
- Ross, L.; Sennyey P. (2008). The library is dead, long live the library! the practice of academic librarianship and the digital revolution. *The Journal of Academic Librarianship*, vol. 34 (2), 145–152. <http://dx.doi.org/10.1016/j.acalib.2007.12.006>
- Sanz-Santacruz, V.; Serrano-Vicente, R.; Iturbide-Tellechea, M.A. (2013). La estancia en bibliotecas extranjeras como instrumento de formación bibliotecaria: la experiencia de la Universidad de Navarra (2003-2010). *Revista Española de Documentación Científica*, 36(1):en001. <http://dx.doi.org/10.3989/redc.2013.1.896>
- Schatzmann, J.; Weilhammer, K.; Stuttle, M.; Young, S. (2006). A Survey of Statistical User Simulation Techniques for Reinforcement-Learning of Dialogue Management Strategies. *Knowledge Engineering Review*, vol. 21(2), 97–126. <http://dx.doi.org/10.1017/S0269888906000944>
- Seneff, S.; Adler, M.; Glass, J.; Sherry, B.; Hazen, T.; Wang, C.; Wu, T. (2007). Exploiting Context Information in Spoken Dialogue Interaction with Mobile Devices. In: Proc. IMUX'07, pp 1–11
- Serrano-Guerrero, J.; Herrera-Viedma, E.; Olivas, J.A.; Cerezo, A.; Romero, F.P. (2011). A Google wave-based fuzzy recommender system to disseminate information in University Digital Libraries 2.0. *Information Sciences*, vol. 181 (9), 1503–1516. <http://dx.doi.org/10.1016/j.ins.2011.01.012>
- Shatte, A.; Holdsworth, J.; Lee, I. (2014). Mobile augmented reality based context-aware library management system. *Expert Systems with Applications*, vol. 41(5), 2174–2185. <http://dx.doi.org/10.1016/j.eswa.2013.09.016>
- Smeaton, A.F.; Callan, J. (2005). Personalisation and recommender systems in digital libraries. *International Journal on Digital Libraries*, vol. 5(4), 299–308. <http://dx.doi.org/10.1007/s00799-004-0100-1>
- Tejeda-Lorente, A.; Porcel, C.; Peis, E.; Sanz, R.; Herrera-Viedma, E. (2014). A quality based recommender system to disseminate information in a university digital library. *Information Sciences*, vol. 261, 52–69. <http://dx.doi.org/10.1016/j.ins.2013.10.036>

Tsuji, K.; Takizawa, N.; Sato, S.; Ikeuchi, U.; Ikeuchi, A.; Yoshikane, F.; Itsumura, H. (2014). Book Recommendation Based on Library Loan Records and Bibliographic Information. *Procedia - Social and Behavioral Sciences*, vol. 147, 478-486. <http://dx.doi.org/10.1016/j.sbspro.2014.07.142>

Vassilakaki, E.; Garoufallou, E. (2013). Multilingual Digital Libraries: A review of issues in system-centered and user-centered studies, information retrieval and user behavior. *The International*

Information & Library Review, vol. 45(1-2), 3-19. <http://dx.doi.org/10.1016/j.iilr.2013.07.002> / <http://dx.doi.org/10.1080/10572317.2013.10766367>

Wu, W.L.; Lu, R.Z.; Duan, J.Y.; Liu, H.; Gao, F.; Chen, Y.Q. (2010). Spoken language understanding using weakly supervised learning. *Computer Speech & Language*, vol. 24(2), 358-382. <http://dx.doi.org/10.1016/j.csl.2009.05.002>