ODISEA: Hacia un entorno inteligente basado en un interfaz en lenguaje natural

Xavier Alamán, Pablo Haya y Germán Montoro
Departamento de Ingeniería Informática
Universidad Autónoma de Madrid, Cantoblanco, 28049-Madrid
{Xavier.Alaman, Pablo.Haya, German.Montoro}@ii.uam.es

Resumen

El proyecto ODISEA propone desarrollar un entorno inteligente con el cual se pueda interactuar en lenguaje natural. Se está implementando un entorno domótico y un entorno ofimático reales, en los que se podrán realizar las tareas habituales en dichos lugares contando con el soporte proactivo del propio entorno. La habitación inteligente tendrá constancia de los ocupantes presentes en cada momento, de sus preferencias y de las tareas que en ese momento están realizando. A partir de esta información del contexto, la habitación inteligente realizará de manera automática acciones que ayuden a los ocupantes e interactuará con ellos mediante diálogos en lenguaje natural.

Desde el punto de vista de la tecnología del habla, se propone la viabilidad de entablar diálogos en lenguaje natural apoyados en una información de contexto muy rica, obtenida a partir de otros modos de interacción (visión artificial y otros tipos de sensores). Se propone diseñar diálogos muy específicos para cada una de las posibles necesidades del usuario, así como un gestor de diálogos que decida cuál de estos se emplea en cada momento, cuándo se cambia de un diálogo a otro, así como qué información se intercambia entre ellos.

Actualmente ya se dispone de un prototipo en funcionamiento, que permitirá avanzar en la definición y el desarrollo de cada uno de los módulos necesarios.

1. Introducción – Estado del arte

Durante los últimos años ha surgido dentro de la comunidad científica que trabaja en el ámbito de las interfaces de usuario (Human Computer Interfaces) una nueva área de trabajo que está recibiendo una gran atención: los así llamados "Entornos Activos" (Active Environments) o "Entornos Inteligentes" (Intelligent Environments, Smart Environments).

Los Entornos Activos interactúan de forma natural con el individuo y le ayudan de manera no intrusiva en la realización de las tareas cotidianas. Los ordenadores dentro de un Entorno Activo quedan ocultos para el usuario; éste obtiene servicios del sistema mediante una interacción en lenguaje natural y sensible al contexto. De este modo se consigue que las habitaciones u oficinas tengan una entidad propia y tomen la iniciativa en la interacción para mejorar la calidad de vida de las personas que las ocupan.

Ejemplos de entornos activos son:

- Intelligent Room [Coen 98], uno de los más desarrollados en la actualidad. Se trata de una Smart Room [Pentland 96] que posee cámaras, micrófonos y otros sensores que intentan interpretar lo que la gente está haciendo con el fin de poder ayudarles del mismo modo que si fueran mayordomos invisibles. Es un proyecto del grupo de investigación del Laboratorio de Inteligencia Artificial del MIT.

- Smart Desks, del mismo departamento, es un tipo de Smart Room, pero especializado en el entorno de negocios. El objetivo de este proyecto es desarrollar un escritorio que actúe como si fuera un ayudante de oficina. Así el escritorio deberá conocer los hábitos, preferencias, y sensaciones del usuario, o simplemente recordar dónde pone las cosas.

- The KidsRoom [Bobick et al 97], también del mismo departamento que el anterior, es un entorno narrativo y de juego interactivo para niños que usa visión computacional y reconocimiento de acciones, de modo que estas informaciones puedan ser utilizadas por el ordenador como datos de entrada para el sistema de control de la narración.

- I.A.E. (Intelligent and Aware Environments) también conocido como Smart Spaces [Essa et al 98], que consiste en espacios que han sido transformados en áreas de trabajo inteligentes donde cámaras, pantallas, micrófonos y otros sensores se fusionan mostrando un entorno integrado.

- EasyLiving [Shafer 99], se trata del esfuerzo investigador de Microsoft para desarrollar Entornos Activos completos, de modo que computadores ubicuos ayuden a las personas en multitud de tareas cotidianas. Uno de los mayores retos de este grupo es que la interacción se realice en lenguaje natural.

2. Hacia un entorno inteligente: el proyecto ODISEA

En el Departamento de Informática de la Universidad Autónoma de Madrid se ha comenzado recientemente un nuevo laboratorio de creación de entornos inteligentes con el objetivo de desarrollar las tecnologías necesarias para llevar a la realidad tales entornos. Ejemplos del tipo de interacciones que se contemplan podrían ser:

- Soporte a la colaboración y la comunicación. Tanto en el ambiente doméstico como en oficinas y otros entornos, el usuario podrá expresar verbalmente instrucciones relativas a necesidades de colaboración o comunicación con otros usuarios, tales como preguntar por la presencia o disponibilidad de otras personas, remitirles recados o mensajes, etc. Los recados o mensajes serán entregados verbalmente, si el destinatario está presente y disponible, o en caso contrario serán convertidos a otros formatos, tales como correo electrónico, mensajes telefónicos, mensajes a un "beeper", etc. El sistema también realizará el seguimiento, cuando sea posible, de la entrega del mensaje, contestación en caso de haberla, e incluso establecimiento de una conversación o diálogo a través de cualquier medio disponible.

- Acceso a la información. El sistema monitorizará y proporcionará información acerca de distintos aspectos del entorno inteligente: aspectos operacionales, tales como el control del estado y funcionamiento de maquinaria de distinto tipo (maquinaria industrial, electrodomésticos, maquinaria de oficina), aspectos relativos a la seguridad (alarmas de incendios, control inteligente de presencia), etc. También se podrá tener acceso verbal y sensible al contexto de fuentes de información disponibles en Internet.

- Seguimiento del entorno ("enviroment awareness"). El sistema ofrecerá de manera espontánea, y según las preferencias de cada usuario, claves y señales no intrusivas acerca del entorno inmediato. Estas señales podrán ser, por ejemplo, pequeñas marcas sonoras que indiquen la llegada de alguien a la casa, que anuncien que los colegas ya se han reunido en la cafetería de la empresa para el desayuno, o que están a punto de cerrar el edificio.

- Control del entorno. El sistema no solo realizará una monitorización pasiva del entorno, sino que permitirá al usuario controlar algunos aspectos de dicho entorno. Este control habitualmente supondrá la interacción con algún tipo de dispositivo o aparato, tal como puede ser una persiana, la luz ambiental, un reproductor de vídeo, un proyector de diapositivas, la calefacción o el aire acondicionado, etc.

Se ha desarrollado un primer prototipo de tal entorno inteligente: el sistema ODISEA. La arquitectura básica de ODISEA descansa sobre dos capas: la capa de interacción con el entorno físico y la capa de contexto. La capa de interacción con el entorno físico incluye el conjunto de dispositivos que componen la infraestructura del entorno inteligente. La capa de contexto ofrece información sobre el estado del entorno, sus ocupantes, las actividades que están realizando y las metas que persiguen, etc. Los módulos de interacción con el usuario, en especial el módulo de diálogos en lenguaje natural, descansan sobre esta última capa.

La capa de contexto

El objetivo del sistema es ofrecer la capacidad de interactuar con el entorno mediante una interfaz en lenguaje natural. Mediante el lenguaje se van a poder controlar los diferentes componentes físicos de la habitación. En sentido inverso, la recolección de información contextual a partir de sensores

multimodales permite realizar un modelo del estado en el cual se encuentra el entorno en cada momento. Esta comunicación en dos direcciones, obliga a tener que considerar un conjunto heterogéneo de dispositivos físicos, cada uno de los cuales con una interfaz distinta de programación. En este sentido, la inclusión de la capa de contexto permite conseguir: por un lado, presentar una única interfaz que abstraiga los detalles de comunicación, por otro poder representar el mundo físico mediante un modelo unificado.

Para implementar esta capa de contexto se utiliza una estructura de datos centralizada, que se denomina pizarra. Esta pizarra constituye un modelo del mundo sobre el que se quiere trabajar. Toda la información relevante al entorno inteligente que se quiere modelizar queda representada en la pizarra, de modo que se crea un "mapa" fácil e intuitivo sobre cómo está compuesto el mundo de entornos con el que se trabaja.

En la pizarra queda representado, además, el flujo de información que se lleva a cabo entre los diferentes componentes físicos que conforman el entorno (micrófonos, altavoces, cámaras, pantallas, etc.) El estado de los diferentes componentes y de su flujo de información es dinámico, adaptándose a las características del entorno de modo que siempre muestren información sobre el estado actual del sistema, información que es fácilmente accesible para poder modificarla tanto por una aplicación software como por un usuario humano.

La incorporación de esta capa de contexto permite que cualquier clase de nuevos componentes exteriores al sistema sea rápida y no presente ningún tipo de complicación, independientemente de las características heterogéneas que puedan tener los diferentes componentes. Para conseguirlo se ha decidido que, tanto para la definición de todos los componentes del entorno como la interfaz con la pizarra, se utilice el lenguaje XML. XML es un lenguaje estándar y de fácil utilización que cuenta, además, con un número cada vez más pujante de analizadores sintácticos y navegadores distribuidos gratuitamente.

Para una primera implementación de la pizarra se ha elegido como punto de partida una herramienta denominada Context Toolkit [Dey, A.K. et al. 1999]. Esta ha sido desarrollada en la Universidad de Georgia Tech. La Context Toolkit se basa en la noción que los desarrolladores de interfaces de usuario tienen del concepto de widget, extrapolándola a un dispositivo físico. De esta forma, se definen widgets reutilizables que encapsulan el comportamiento de uno o varios dispositivos.

La pizarra internamente queda constituida por un conjunto de widgets: cada uno conserva el estado y el comportamiento de un dispositivo físico. De cara al exterior, la pizarra es accesible por otros componentes software a través de un único puerto de comunicación. En un futuro próximo a partir de la definición de la pizarra mediante XML se tendría que obtener de forma automática la implementación de todos los widget.

La capa de interacción con el entorno físico

Dentro del laboratorio de pruebas, existen numerosos dispositivos y equipos (sensores, actuadores, cámaras, micrófonos, video-proyectores, reproductores de vídeo, etc.). Cada uno de los componentes se conecta a una red física distinta. En principio, gracias a la capa de contexto, cualquier tipo red que implemente una pasarela a nuestra capa de contexto puede integrarse dentro de la capa de interacción con el entorno físico.

Considerando un enfoque más práctico se ha decidido emplear dos tecnologías como punto de partida de la construcción del primer prototipo. En la parte de control de dispositivos, sensores (de presencia, temperatura, luminosidad...) y actuadores (interruptores, relés, controladores de motores), se utiliza el bus europeo EIB [http://idaho.eiba.com/home.nsf], que pretende crear un estándar para este campo dentro de la Unión Europea. Este bus permite integrar componentes domóticos y programarlos de forma rápida y sencilla. El bus EIB ya ha sido instalado con éxito en diversos edificios de índole público y privado, controlando, desde pequeñas habitaciones hasta edificios enteros, dentro del marco de la Unión Europea. En la parte del flujo continuo de información (micrófonos y videocámaras), se ha dispuesto una red Ethernet que conecta los diferentes dispositivos.

3. Interacción en lenguaje natural dentro del entorno inteligente

En ODISEA, los usuarios se comunican con el sistema mediante el habla, utilizando expresiones no restringidas en lenguaje natural. A su vez, el entorno responde al usuario utilizando expresiones

adecuadas. Dentro del entorno inteligente se pretende que las personas mantengan diálogos robustos en lenguaje natural. Por diálogo robusto se entiende aquel en el que el usuario del sistema puede emplear expresiones libres y generales (y por tanto no restringidas a un lenguaje predeterminado de comandos) para acceder a las funciones y servicios del entorno.

Para el proceso de reconocimiento de la voz se emplea la herramienta comercial ViaVoice de IBM. Esta es la herramienta comercial más extendida en la actualidad para reconocimiento de voz y también se utiliza en otros proyectos de similares características (p. ej. [Coen 98]). La comunicación con el reconocedor se realiza utilizando la Java Speech API, lo que garantiza que se puede cambiar en cualquier momento el motor de reconocimiento sin que esto afecte a la implementación de la aplicación. En cuanto a la síntesis de voz se efectúa utilizando la capacidad *Text to Speech* de ViaVoice. Las voces se pueden modular para que el usuario perciba mayor sensación de naturalidad.

En segundo lugar, para abordar el problema de hacer que los diálogos sean libres y sin restricciones se están utilizando tres factores fundamentales:

1. El usuario se encuentra dentro de un entorno restringido que permite poder "intuir" qué tipo de comunicación establecerá con el sistema. Si el habitante del entorno se encuentra en una oficina, no tiene sentido esperar que pregunte por el precio de los tomates, y si se encuentra en un supermercado no tiene sentido procesar la información correspondiente a la pregunta de si ya ha llegado su jefe.

2. Aunque el usuario puede establecer múltiples conversaciones con el sistema dentro de un mismo entorno, estas conversaciones están dirigidas por el sistema y el usuario siempre se encuentra dentro de algunos de los diálogos que tiene previstos el sistema. Por lo tanto, si el usuario se encuentra dentro de un diálogo cuyo fin es enviar un correo electrónico, el sistema reconoce esta situación, y resulta previsible que las próximas expresiones que realice tengan como objetivo terminar de realizar la tarea de enviar el mensaje.

3. El sistema cuenta con una información de contexto muy rica que se encuentra en la pizarra descrita anteriormente. La información de contexto referente al entorno y a los diálogos que se están produciendo se almacena en la pizarra para que pueda ser utilizada por los diferentes módulos que componen el sistema, incluido el módulo de lenguaje natural. La pizarra almacena por ejemplo información sobre el estado de las luces o el último elemento del entorno al que se refirió el usuario. Si el usuario solicita al entorno que le diga la hora y a continuación le pide que le repita la información, el sistema sabrá que el usuario desea volver a oír qué hora es.

Comentario: Poner número de apartado de arquitectura

Para que el módulo de diálogos pueda procesar la información que se recibe del reconocedor se utilizan un conjunto de analizadores y gramáticas integradas en las herramientas MACO+ y Relax [Carmona, J. et al 1998] y Tacat [Atserias, J. et al. 1998]. Estas herramientas y recursos lingüísticos han sido cedidos por el Grupo de Investigación en Procesamiento del Lenguaje Natural de la Universidad Politécnica de Catalunya y el Laboratorio de Lingüística Computacional de la Universidad de Barcelona.

Se empieza realizando un análisis morfológico utilizando MACO+. Esta herramienta produce etiquetas morfosintácticas para cada palabra de la oración. A continuación se utiliza el desambiguador morfosintáctico Relax capaz de desambiguar palabras dado un conjunto de restricciones. Por último, se emplea el analizador sintáctico Tacat, que realiza análisis de texto sin restricciones.

Una vez obtenido un análisis sintáctico de la expresión que el usuario ha pronunciado, se analiza el resultado para determinar qué ha dicho el usuario. Para ello puede utilizar la información almacenada en la pizarra. Por ejemplo, si el usuario quiere que se enciendan las luces del entorno, en un caso simple, el sistema puede esperar una expresión que contenga un grupo verbal con el verbo encender y un sintagma nominal que contenga el sustantivo luz. Estudiemos tres posibles casos para este ejemplo:

- Si el sistema reconoce el verbo y el sustantivo de forma correcta, puede utilizar la información de la pizarra para saber si la luz ya está encendida y responder al usuario o simplemente encenderla.

- También puede ocurrir que el sistema no hubiera sido capaz de entender el sustantivo luz y sólo el verbo encender. En ese caso, se puede utilizar la información de la pizarra para determinar qué

elementos del entorno son susceptibles a ser encendidos y preguntar al usuario sobre cuál quiere actuar.

- Por último el sistema puede entender el sustantivo luz pero no qué acción se debe realizar sobre ella. Para procesar esta información puede consultar la pizarra y determinar que, por ejemplo, si la luz se encuentra apagada en esos momentos y la información almacenada sobre el sensor fotosensible indica un entorno poco luminoso, el usuario probablemente querrá se enciendan las luces.

El sistema consta inicialmente de un modelo de gestión de diálogos muy simple que irá aumentando en complejidad según avance el proyecto. Los diferentes diálogos que el entorno es capaz de soportar son independientes entre sí, por lo que resulta sencillo quitar, añadir o modificar un diálogo sin afectar al resto del entorno. Puede haber, por lo tanto, un diálogo encargado del control de las luces, otro sobre la presencia de personas en alguna otra habitación del entorno, etc. En cualquier momento se puede añadir uno nuevo, por ejemplo, sobre el manejo del correo electrónico. Estos diálogos se basan en la idea de guiones de conversación y plantillas de tarea:

- Los guiones de conversación se componen de una secuencia de patrones de palabras, con sus respectivas relaciones, y de sus respuestas correspondientes. Las frases pronunciadas por el usuario, y adecuadamente analizadas, se comparan con la colección de patrones de palabras almacenadas en el guión de la conversación. Cuando se produce una coincidencia se ejecuta la parte del guión correspondiente, lo que puede desencadenar la realización de ciertas acciones o de una respuesta al usuario. Para ello también se utiliza la información de contexto almacenada en la pizarra.

- Las plantillas de tarea definen los parámetros requeridos para completar una determinada tarea. Cuando se produce una coincidencia con un patrón, el guión de conversación correspondiente puede invocar el procesamiento de la frase con respecto a una plantilla de tarea específica hasta que ésta se complete.

Todos los diálogos que soporta el entorno discurren en paralelo y están controlados de forma centralizada por el supervisor de diálogos. Su funcionamiento sigue los siguientes pasos:

1. El supervisor de diálogos se encarga de recibir el texto del reconocedor y analizarlo para, a continuación, enviar esta información a todos los diálogos que posee el entorno.

2. Cada diálogo compara el texto etiquetado con sus guiones de conversación y responde al supervisor si considera que la expresión pronunciada por el usuario pertenece a su diálogo.

3. El supervisor de diálogos da prioridad al diálogo que se encontraba activo en ese momento o, en su defecto, da paso al siguiente diálogo que ha optado por el turno.

4. El diálogo que recibe el control por parte del supervisor es el que se encarga de ejecutar las acciones descritas en su guión de conversación.

Con este método se pretende conseguir que resulte sencillo crear nuevos diálogos y añadirlos o modificar y mejorar los que ya existen. El usuario puede saltar de un diálogo a otro y el sistema sigue la pista de qué acción se pretende realizar en cada momento. Si en mitad de un diálogo sobre el correo electrónico el usuario pidiera hacer una llamada de teléfono, el sistema podría cambiar al diálogo sobre realizar llamadas telefónicas para luego volver al diálogo sobre el correo electrónico o volver a cambiar a otro.

El gestor de diálogos desarrollado dentro del proyecto ODISEA se encuentra en sus primeras fases de su desarrollo. Los módulos aquí expuestos se encuentran implementados, en proceso de construcción o se espera que funcionen en un breve plazo de tiempo.

4. Conclusiones y trabajo futuro

El primer intento de plasmar nuestras ideas en la realidad ha dado como fruto el prototipo que se describe a continuación. El sistema implementado se compone de las tres capas mencionadas en los apartados

anteriores. La capa de interacción con el entorno físico la componen, en la parte de control, interruptores, relés para controlar las luces de la habitación, un sensor de presencia y un display, todos ellos conectados al bus EIB. Para la parte de flujo de información continua se han dispuesto un micrófono y dos pares de altavoces, cada uno conectados a un ordenador distinto, formando los tres ordenadores una red local. Se dispone de un tercer ordenador que hace de pasarela entre el bus EIB y la red de ordenadores. Los programas ViaVoice, MACO+, Relax y Tacat residen en el mismo ordenador al cual se conecta el micrófono.

Se ha definido un conjunto restringido de diálogos que permiten interactuar con los distintos dispositivos del bus EIB (encender/apagar las luces, poder desplegar mensajes en el display, preguntar si hay alguien en la habitación). En relación con las capacidades de actuación sobre los flujos de información, por ejemplo, se ha desarrollado un diálogo que permite cambiar la salida de audio de uno de los altavoces a otro.

La capa de contexto se compone de un widget para cada uno de los dispositivos físicos (luces, sensor de presencia, relé, altavoces), así como de información de contexto sobre los usuarios presentes. Los widgets se distribuyen en los diferentes ordenadores utilizando cada uno puerto de comunicación para recibir y enviar mensajes. Se ha desarrollado una interfaz que permite al módulo de diálogo acceder a los diferentes widget utilizando un único puerto de comunicación.

El prototipo está en funcionamiento, y va a permitir en el futuro investigar en diversas áreas de interés relacionadas con las tecnologías del habla. El objetivo es implementar este prototipo en los despachos de varios de los participantes en el proyecto, y empezar a emplearlo como herramienta de trabajo diario.

Referencias

ABOWD, G.D.; ATKESON, C.; FEINSTEIN, A.; GOOLAMABBAS, Y.; HMELO, C.; REGISTER, S.; SAWHNEY, N. and TANI, M. 1996. Classroom 2000: Enhancing classroom interaction and review. *Technical Report GIT-GVU-96-21*, GVU Center, Georgia Institute of Technology.

ATSERIAS, J.; CARMONA, J.; CASTELLÓN, I.; CERVELL, S.; CIVIT, M.; MÁRQUEZ, L.; MARTÍ, M.A.; PADRÓ, L.; PLACER, R.; RODRÍGUEZ, H.; TAULÉ, M. and TURMO, J. 1998. Morphosyntactic Analysis and Parsing of Unrestricted Spanish Text, In *Proceedings of the 1st International Conference on Language Resources and Evaluation* (LREC'98, Granada, Spain).

BOBICK, A.; INTILLE, S.; DAVIS, J.; BAIRD, F.; PINHANEZ, C.; CAMPBELL, L.; IVANOV, Y.; SCHÜTTE, A. and WILSON, A. 1997 The KidsRoom: A Perceptually-Based Interactive and Immersive Story Environment, *Technical report 398*, MIT Media Laboratory, Perceptual Computing Section.

CARMONA, J.; CERVELL, S.; ATSERIAS, J.; CERVELL S.; MÁRQUEZ, L.; MARTÍ, M.A.; PADRÓ, L.; PLACER, R.; RODRÍGUEZ, H.; TAULÉ, M. and TURMO, J. 1998. An Environment for Morphosyntactic Processing of Unrestricted Spanish Text. In *Proceedings of 1st International Conference on Language Resources and Evaluation* (LREC'98, Granada, Spain).

COEN, M.H. 1998. Design Principles for Intelligent Environments. In *Proceedings of the AAAI Spring Symposium on Intelligent Environments* (AAAI98).

DEY, A.K.; ABOWD, G.D. and SALBER, D. 1999. A context-based infrastructure for smart environments. In *Proceedings of 1st International Workshop on Managing Interactions in Smart Environments* (MANSE'99, Dublin, Dec 1999), P. Nixon, G. Lacey and S. Dobson, Eds. Springer-Verlag, London, 114-128.

ESSA I., ABOWD G. Y ATLESON C. "Ubiqiitous Smart Spaces", A white paper submitted to DARPA, February 1998.

PENTLAND, A. 1996. Smart rooms. *Scientific American*, 274, 4, 68-76.

SHAFER, S. "Ten dimensions of ubiquitous computing". In *Proceedgins of 1st International Workshop on Managing Interactions in Smart Environments (MANSE'99)*, December, 1999.