

# Propuesta de definición de una interfaz dinámica para la interacción con el entorno

Germán Montoro, Pablo A. Haya y Xavier Alamán

Escuela Politécnica Superior  
Universidad Autónoma de Madrid  
28049 Madrid

{German.Montoro, Pablo.Haya, Xavier.Alaman}@uam.es

## Resumen

Los entornos inteligentes requieren de nuevas interfaces que permitan una interacción natural. El desarrollo de estas interfaces se ha de realizar de forma estándar, considerando las características dinámicas de estos entornos. En esta comunicación se presentan los resultados en el desarrollo de un lenguaje de descripción para la creación automática de una interfaz de diálogos orales que se adapta a las características del entorno inteligente dado.

## 1. Introducción

Los entornos inteligentes [3] han aparecido como un nuevo campo de investigación dentro del área de interfaces de usuario. Estos entornos proporcionan nuevas posibilidades de interacción [10], ofreciendo nuevos retos a los diseñadores de las interfaces [8]. Los ordenadores y dispositivos computacionales quedan ocultos a los usuarios y éstos pueden obtener los servicios del sistema, por ejemplo, mediante interfaces orales en lenguaje natural sensibles al contexto.

La aparición de estos entornos hace necesario construir nuevas interfaces que permitan interactuar con los usuarios de forma natural. Estas interfaces pueden ser capaces de entablar conversaciones relativas al entorno, sus elementos y los servicios que puede proporcionar a sus usuarios. Las interfaces de diálogos orales se han de adaptar a estos sistemas [5, 7], de modo que puedan hacer frente a los nuevos retos que proporcionan.

En esta comunicación presentamos los resultados en el proceso de definición de una interfaz de diálogos orales para entornos

inteligentes que se adapta al dominio de cada entorno. Los diálogos se crean de forma automática y permiten interactuar con el entorno y controlar sus dispositivos mediante interacción oral en lenguaje natural.

Para llevar a cabo nuestra investigación hemos construido un entorno inteligente real. Este entorno está formado por un laboratorio amueblado como una sala de estar y en el que se ha dispuesto un conjunto de dispositivos. Entre ellos se encuentran luces y controles de iluminación, un mecanismo de apertura de la puerta de entrada, un detector de presencia, tarjetas de acceso personales, altavoces, micrófonos, un sintonizador de radio, cámaras IP, etc.

Dadas las características heterogéneas y dinámicas de los entornos inteligentes resulta fundamental que los diálogos de la interfaz se puedan crear automáticamente, adaptándose a las características específicas de cada entorno. Para ello, junto a la definición del entorno y sus entidades, se define la información lingüística de cada una de ellas, la descripción del entorno y las gramáticas asociadas. Estos elementos se emplean para crear la interfaz de forma automática y adaptada a cada espacio.

El sistema de diálogos creado se basa en una estructura en árbol, donde las partes lingüísticas están representadas en nodos. Además presenta un conjunto de gramáticas creadas automáticamente que definen todas las posibles formas de interacción dentro del entorno.

Este árbol se utiliza en los procesos de comprensión e interacción con los usuarios, esto es, para realizar interpretación y generación de oraciones [4].

```

<classes>
  definición_clase
  [, definición_clase ] ...
</classes>
definición_clase:
<class name="nombre" extends="tipo">
  <property name="nombre_propiedad">
    [conjunto_propiedades, ...]
  </property>
  [, <property name="nombre_propiedad">
    [conjunto_propiedades, ...]
  </property>
  ] ...
</class>
conjunto_propiedades:
<paramSet name="nombre_conjunto">
  <param name="nombre_parámetro">valor</param>
  [, <param name="nombre_parámetro">valor</param> ] ...
</paramSet>

```

Figura 1. Sintaxis del Documento de Definición de las Clases de Entidad (DDCE)

## 2. Lenguajes de definición de interfaces de usuario

XML permanece como una solución para la estandarización de la interoperabilidad entre aplicaciones. Es por esto que se siguen empleando nuevos lenguajes basados en XML para definir interfaces de usuario (XML-UIDL). Estos lenguajes tienen la ventaja de ser transparentes a diferentes tecnologías de interfaz y de proporcionar un recurso homogéneo para modos de comunicación heterogéneos [1].

Estos lenguajes XML empleados para la representación de interfaces se deben poder aplicar a cualquier objetivo, en cualquier contexto y debe ser personalizables, flexibles y extensibles [9]. A su vez, deberían separar los elementos de la interfaz de su presentación. Los elementos de la interfaz de usuarios se deben representar de forma explícita en un formato que se pudiera trasladar a cualquier contexto posible.

Dos lenguajes representativos en este campo son:

- UIML [2], se trata de un lenguaje XML que permite la creación de interfaces de usuario para cualquier dispositivo, lenguaje destino y sistema operativo. Describe la apariencia de la

interfaz, la interacción del usuario con la misma y cómo está conectada con la lógica de la aplicación.

- XIML [6], un lenguaje de representación de la interfaz basado en XML que proporciona un apoyo de la funcionalidad a través del ciclo de vida de una interfaz de usuario: diseño, desarrollo, operación, gestión, organización y evaluación

El lenguaje desarrollado para la definición de las interfaces en los entornos inteligentes utiliza algunas de las características presentes en este tipo de lenguajes de descripción, incorporando nuevas funcionalidades que se adaptan a las peculiaridades de los entornos para los que desarrollan la interfaz.

## 3. Definición de las entidades de un entorno inteligente

Los tipos posibles de entidades que pueden estar presentes en un entorno inteligente se definen en uno o varios documentos de descripción de las clases de entidad (DDCE). Cada clase de entidad define las características comunes que tiene toda entidad de ese tipo. Las instancias de esa clase de entidad heredan las características comunes de la clase y pueden, en caso de ser necesario, añadir

propiedades específicas de la instancia (ver siguiente apartado).

De este modo, cuando se diseña un nuevo tipo de dispositivo físico o de aplicación en el entorno se acompaña de un DDCE, que corresponde con la documentación en XML de la descripción de las características comunes de ese tipo de entidad.

Cada clase se define mediante sus propiedades (especificadas mediante la etiqueta *property*) y, opcionalmente, un conjunto de parámetros comunes (representado con las etiquetas *paramSet* y *param*). Los parámetros pueden estar asociados a una propiedad concreta o a toda la clase. Además, cada clase hereda las propiedades de uno de los tipos al que pertenece (mediante el atributo *extends*). Estos tipos definen características comunes. La figura 1 muestra la sintaxis de un DDCE.

Existen varios tipos comunes de clase, como por ejemplo, *room*, *device*, *person*. Toda clase ha de heredar de uno de estos tipos o de una clase que haya heredado de ellos. La definición de un tipo común contiene las características básicas que pueden compartir todas las clases del mismo tipo. Los tipos comunes de clases vienen dados por el sistema y no pueden ser modificados por los usuarios. Los usuarios pueden definir nuevas clases basándose en estos tipos.

#### 4. Definición de la interfaz

De la misma manera que ocurre con otros elementos, la definición de la información lingüística asociada a las entidades se establece en unos documentos denominados Documentos de

Parametrización de las Clases de Entidad (DPCE), relacionándola con las clases de entidad. De este modo las instancias de esos tipos heredan automáticamente todas las propiedades definidas.

La definición de la interfaz se compone de dos fases diferenciadas que se pueden realizar en momentos diferentes y por personas distintas:

- En primer lugar se deberá definir en los DPCE la información relacionada con cada nueva clase de entidad. Esta comprende la información lingüística de interacción, los métodos necesarios para la automatización del sistema y, opcionalmente, de nuevas plantillas de gramática. Esta información se definirá una única vez y será compartida por todas las entidades del mismo tipo.
- En segundo lugar se deben definir en unos Documentos de Definición de las Entidades del Entorno (DDEE) qué entidades se encuentran presentes en el entorno y de qué tipo son. La información lingüística de cada entidad del entorno vendrá dada por la que han establecido anteriormente las clases de entidad en los DPCE. De este modo es posible que para crear una interfaz de diálogos orales de interacción con el entorno sólo sea necesario definir los elementos que se encuentran en el mismo, sin necesidad de modificar o añadir ningún tipo de información lingüística adicional. Sin embargo, la información que hereda cada entidad también se puede personalizar, adaptándola a las características especiales del entorno.

```

<classes>
  definición_clase
  [, definición_clase ] ...
</classes>
definición_clase:
<class name="nombre">
  <property name="propiedad">
    <paramSet name="dialogue">
      <param name="action1"> Nombre_de_la_acción </param>
      <param name="skeleton1"> Parte Palabra [, Parte Palabra ]...
    </param>
    [, <param name="skeleton2"> Parte Palabra
      [, Parte Palabra ]...
    </param> ] ...
    [, <param name="action2"> Nombre_de_la_acción </param>
      <param name="skeletonn"> Parte Palabra
      [, Parte Palabra ]...
    </param>
    [, <param name="skeletonm"> Parte Palabra
      [, Parte Palabra ]...
    </param> ] ... ]...
    </paramSet>
  </property>
</class>
Parte:
VP | OP | LP | IOP | MODALP | QP | MP

```

Figura 2. Sintaxis de la información lingüística que se adjunta en el DPCE

## 5. Definición de la información lingüística asociada a las clases de entidad

Como se acaba de mencionar, cada clase de entidad tiene asociada su propia información lingüística que establece todas las posibles interacciones que se pueden realizar con las entidades de ese tipo. Esta información se clasifica en siete posibles partes lingüísticas, sin perjuicio de poder ser ampliadas fácilmente añadiendo nuevas partes. Las siete partes lingüísticas que se proponen son:

- Parte verbal (VP). Describe las acciones que se pueden llevar a cabo con la entidad.
- Parte objeto (OP). Establece los posibles nombres que pueden tomar las entidades que reciben la acción del verbo.
- Parte de ubicación (LP). Describe su situación física dentro del entorno.

- Parte de objeto indirecto (IOP). Especifica a quién o a qué va dirigida la acción que se realiza.
- Parte modal (MODALP). Indica el modo en que se debe realizar la acción.
- Parte cuantificadora (QP). Define un valor o cantidad que se aplica sobre la acción que se realiza.
- Parte modificadora (MP). Añade información calificativa a alguna de las partes anteriores.

La información lingüística asociada a las clases de entidades se define mediante los DPCE. Esta información se compartirá posteriormente por todas las entidades de ese tipo que se definan en el entorno. Esto es, dos entidades del mismo tipo heredan la misma información lingüística que posteriormente puede ser personalizada, dependiendo de las características propias del entorno (ver siguiente apartado).

```

<instances>
  definición_instancia
  [ , definición_instancia ] ...
</instances>

definición_instancia:
<entity name="nombre" type="clase_entidad"/>
[ , <entity name="nombre" type="clase_entidad"/> ] ...

```

Figura 3. Sintaxis del Documento de Definición de las Entidades del Entorno (DDEE)

Para añadir la información referente a la interfaz lingüística se ha de adjuntar un conjunto de parámetros bajo la etiqueta de nombre *dialogue*. A continuación habrá un parámetro *action* por cada una de las posibles acciones que se pueden realizar con la entidad y, para cada uno de ellos uno o varios parámetros *skeleton* que definen los posibles esqueletos de oraciones que se pueden invocar para realizar esa acción. La sintaxis se muestra en la figura 2.

Los parámetros *action* identifican una nueva acción y contienen la descripción del tipo de acción que se realiza. Los esqueletos de oraciones se identifican mediante el parámetro *skeleton* y contienen las palabras claves que los constituyen y que permiten realizar la acción que tienen asociada. Cada palabra clave debe ir precedida por la parte lingüística que representa (*VP*, *OP*, *LP*, *IOP*, *MODALP*, *QP* o *MP*). Si se desea especificar sinónimos de una misma parte lingüística se pueden escribir dos o más palabras clave seguidas. Los esqueletos de oraciones pueden iniciarse con cualquier parte lingüística y éstas se pueden repetir cuantas veces sea necesario.

Este documento se puede editar y modificar para adaptarlo a distintas formas orales de interacción que se pueden presentar. Por ejemplo, se podrían añadir, quitar o cambiar algunos de los sinónimos o de las partes para adaptar la interacción a modos regionales o transnacionales de interaccionar, nombrar o dirigirse a las entidades. Se ha buscado que los modos de interacción sean fácilmente definibles y reconfigurables, pudiendo cambiar, crecer o adaptarse a nuevas necesidades fácilmente. No es necesario modificar la implementación de la interfaz de diálogos, sino que basta con editar y modificar las definiciones establecidas en el DPCE para que los cambios estén presentes en la nueva interfaz.

## 6. Definición de las instancias de entidad

Una vez definida la información lingüística asociada las clases de entidades, y dado que todas las entidades del mismo tipo comparten las mismas posibilidades de interacción, en muchas ocasiones basta con definir qué entidades se encuentran en el entorno para que automáticamente se cree una interfaz oral adaptada al mismo.

La definición de los elementos que se encuentran en el entorno se especifica en los DDEE. La figura 3 muestra la sintaxis de la definición de las entidades.

Sin embargo, en ocasiones será necesario (o simplemente recomendable) especificar información lingüística relativa al entorno concreto sobre el que se crea la interfaz. Este es el caso, por ejemplo, de un entorno donde aparezcan varias entidades del mismo tipo, en donde será necesario añadir nueva información que permita diferenciar entre ellas. También se produce esta circunstancia cuando existen características particulares de la entidad en el entorno, que no se pueden especificar en una definición general del tipo de entidad (tal es el caso del color, tamaño, posición, etc.).

Para solventar estas circunstancias, además de la simple definición de las entidades que están presentes en el entorno, se utilizan los Documentos de Parametrización de las Entidades del Entorno (DPPE), que permiten especificar nueva información lingüística concreta para cada entidad. Esto se realiza empleando dentro del conjunto *dialogue* el parámetro *add* en la propiedad donde se quiere añadir la nueva información. A continuación, el atributo debe ir seguido por un número, que señala un esqueleto de oración concreto al que añadir la nueva información o por la palabra *all*, que especifica

```

<entity name="nombre">
  <property name="propiedad">
    <paramSet name="dialogue">

      <param name="add_(1, all)">valor</param>
      [ , <param name="add_2">valor</param> ] ...

    </paramSet>
  </property>
</entity>

```

Figura 4. Sintaxis de la parametrización lingüística en un DPEE

que la información se debe añadir a todos esqueletos de oración de esa propiedad. La figura 4 muestra la sintaxis de la parametrización lingüística de la propiedad de una entidad.

Nuevamente, la idea fundamental es permitir configurar, describir, modificar y adaptar de forma sencilla las interacciones lingüísticas que se pueden llevar a cabo con las entidades del entorno. Una vez definida en los DPCE la información lingüística asociada a las clases de entidades (posiblemente por otra u otras personas) un diseñador de la interfaz del entorno sólo necesita definir mediante los DDEE qué entidades están presentes en el entorno. Este u otros diseñadores especializados podrán a su vez, por medio de los DPEE, modificar estas definiciones, adaptando la interfaz a las características propias de interacción del entorno determinado.

Toda esta información se une en el Documento de Descripción del Entorno (DDE), que sirve de soporte para la creación automática de la interfaz de diálogos orales.

La definición de la información lingüística se realiza de forma análoga a la que permite definir las entidades, sus parámetros y otras interfaces. De este modo se consigue una forma estándar y homogénea de definición del entorno, sus elementos y posibilidades multimodales (por ejemplo, mediante interfaz Web y de diálogos orales) de interacción.

## 7. Creación automática de la interfaz de diálogos orales

Con la información lingüística obtenida del DDE se crea de forma automática la interfaz de diálogos orales adaptada a las características concretas del entorno.

El proceso de creación se basa en dos pasos que se realizan de forma paralela. El primero consiste en la creación de un conjunto de gramáticas adecuadas para la interacción con el entorno. El segundo se basa en la construcción de un árbol lingüístico que será utilizado en los procesos de interpretación y generación.

En el proceso de creación de la interfaz el sistema lee el DDE. Para cada una de las entidades representadas en el documento (esto es, aquellas que están presentes en el entorno) se obtiene la información lingüística asociada a la clase de entidad y, si es necesario, se añade la nueva información lingüística representada en el DDE que es específica para esa entidad. Empleando esta información se construyen las gramáticas y el árbol de representación lingüística.

Al mismo tiempo que se crean las gramáticas de interacción con las entidades del entorno se construye automáticamente un árbol lingüístico, que será un elemento fundamental en los procesos de interpretación y generación.

El procedimiento de construcción parte de un nodo raíz vacío. Cada una de las partes de los esqueletos de oraciones asociados a las entidades se convierte en información que se añade al árbol, bien sea en forma de un nuevo nodo o completando la información de un nodo existente. Cada nuevo nodo se compone de:

- La palabra y el tipo de parte especificados en el esqueleto de la oración.
- La lista de entidades de la pizarra que contiene esa palabra a ese mismo nivel.
- Información sobre si, para alguna entidad, esa palabra inicia un subdiálogo o no.
- Información que determina si el nodo se encuentra en un estado habilitado o deshabilitado.

```

<class name="dimmerlight">
  <property name="status">
    <paramSet name="dialogue">
      <param name="action1">encender</param>
      <param name="skeleton1"> VP encender poner
      OP luz NCF000:lámpara MP ambiente halógena </param>
      <param name="action2"> apagar </param>
      <param name="skeleton2"> VP apagar quitar
      OP luz NCF000:lámpara MP ambiente halógena </param>
    </paramSet>
  </property>
  <property name="value">
    <paramSet name="dialogue">
      <param name="action1">subir</param>
      <param name="skeleton1"> VP subir aumentar
      OP luz NCF000:lámpara MP ambiente halógena</param>
      <param name="skeleton2"> VP subir aumentar OP intensidad
      LP luz NCF000:lámpara MP ambiente halógena </param>
      <param name="action2"> bajar </param>
      <param name="skeleton3"> VP bajar disminuir reducir
      OP luz NCF000:lámpara MP ambiente halógena </param>
      <param name="skeleton4"> VP bajar disminuir reducir
      OP intensidad LP luz NCF000:lámpara
      MP ambiente halógena </param>
    </paramSet>
  </property>
</class>

```

Figura 5. Información lingüística para una entidad de tipo luz regulable utilizada en el entorno

- La lista de los nombres de las acciones asociadas a cada entidad (este nombre viene dado por el valor de los parámetros *action* en la definición de la información lingüística de las clases de entidad.

Cuando el sistema se encuentra una parte donde aparecen varios sinónimos crea un nodo con cada uno de éstos. A partir de ese momento, los restantes hijos de ese mismo esqueleto de oración colgarán de cada uno de los sinónimos.

### 8. La interfaz de diálogos orales en el entorno inteligente implementado

Dentro del entorno inteligente implementado se han definido varias interfaces que permiten interactuar con las entidades que componen al mismo.

Una de ellas consiste en una interfaz de diálogos orales que permite interactuar con las cinco luces del entorno, la cerradura electrónica

de la puerta de entrada y un sintonizador de radio con catorce emisoras diferentes. Además se ha simulado la presencia de un aparato de aire acondicionado y de un sistema sencillo para realizar llamadas de teléfono a cuatro posibles personas. Por último existe un módulo que no está asociado a ninguna entidad del entorno que permite establecer saludos, despedidas, agradecimientos, etc. Estos elementos corresponden con siete tipos diferentes de entidades: luz regulable, luz fluorescente, luz de foco, puerta, radio, aire acondicionado y teléfono más uno de buenas maneras.

Para cada uno de los tipos de entidad se ha definido su información lingüística. Un ejemplo de la definición lingüística que se ha utilizado en la interfaz empleada en el entorno real se presenta en la figura 5, donde se muestra la definición lingüística de una clase *dimmerlight* y en la figura 6 donde se representa la parametrización de la instancia *lampv1* de tipo *dimmerlight* empleada en el entorno desarrollado.

```

<entity name="lampv1">
  <property name="status">
    <paramSet name="dialogue">
      <param name="add_all">LP izquierda</param>
    </paramSet>
  </property>
  <property name="value">
    <paramSet name="dialogue">
      <param name="add_all">LP izquierda</param>
    </paramSet>
  </property>
</entity>

```

Figura 6. Definición de los parámetros de la instancia lampv1 empleada en el entorno

Una vez definida la información lingüística para cada tipo de entidad sólo es necesario definir qué entidades se encuentran presentes en el entorno.

Con estos elementos se crean de forma automática ocho gramáticas distintas, una por cada tipo de entidad. A continuación se construye el árbol lingüístico adaptado al entorno implementado. Este árbol se compone de un total de 310 nodos diferentes. Las combinaciones de caminos existentes en el árbol, la posibilidad de saltar nodos del árbol y el empleo de las gramáticas creadas hacen que las posibles interacciones que la interfaz puede establecer se eleven de forma considerable, permitiendo interacciones múltiples y naturales con las entidades del entorno.

### Agradecimientos

Este trabajo está financiado por el Ministerio de Educación y Ciencia, número de proyecto TIN2004-03140.

### Referencias

- [1] Abrams, M., Phanouriou, C., Batongbacal, A.L., Williams, S., and Shuster, J.E. UIML: An Appliance-Independent XML User Interface Language. In *Proceedings of the Eighth International WWW Conference*, Toronto, Canada, 1999.
- [2] Ali, M.A., Pérez-Quiñones, M.A., Abrams, M., and Shell, E. Building Multi-Platform

User Interfaces with UIML. In *Proceedings of CADUI*, 2002.

- [3] Coen, M.H. Design Principles for Intelligent Environments. In *Proceedings of the AAAI Spring Symposium on Intelligent Environments*, Palo Alto, California, 1998.
- [4] Montoro, G., Alamán, X. and Haya, Pablo A. 2004. Interacción con entornos inteligentes mediante diálogos basados en contexto. *Interacción* 2004.
- [5] Paternò, F., and Santoro, C. One Model, Many Interfaces. In *Proceedings of CADUI*, 2002.
- [6] Puerta, A. and Eisenstein, J. XIML: A Universal Language for User Interfaces. *White paper*. Disponible en <http://www.xml.org/Docs.asp>. 2001.
- [7] Rayner, M., Lewin, I., Gorrell, G., and Boye, J. Plug and Play Speech Understanding. *2nd SIGdial Workshop on Discourse and Dialogue*, September 2001.
- [8] Shafer, S., Brumitt, B., and Cadiz, JJ. Interaction Issues in Context-Aware Intelligent Environments. *Human-Computer Interaction*, 16, 363-378, 2001.
- [9] Trewin, S., Zimmermann, G., and Vanderheiden, G. Abstract user interface representations: How well do they support universal access?. In *Proceedings of the 2nd ACM International Conference on Universal Usability*, Vancouver, Canada, 2003.
- [10] Weiser, M. The world is not a desktop. *ACM Interactions*, 1, 1, 7-8, 1994.