Estándar VoiceXML

Víctor Álvarez García
victoralvarez@uniovi.es

Resumen

El W3C lidera desde Octubre de 1994 el desarrollo de la Web, siendo uno de los objetivos básicos del consorcio promover la interoperabilidad entre las distintas tecnologías que tienen cabida y forman Internet.

Una de estas tecnologías son las aplicaciones de voz o sistemas de diálogo, para los cuales existen varias propuestas de estandarización entre las que podemos encontrar VoiceXML[1], un lenguaje de etiquetas basado en XML cuyas características e impacto actual y futuro son analizados en esta ponencia.

La actual proliferación y perspectiva de incremento de los sistemas de diálogo, remarcan la importancia de este estudio, que toma como partida la especificación del W3C, analiza la aportación de VoiceXML[1] y finaliza con la mención de algunas herramientas que incluyen este y otros estándares de voz entre sus características.

Palabras clave

VoiceXML, sistemas de diálogo, estándares, W3C, VoIP, IVR, PBX, Asterisk

1 El estándar VoiceXML

VoiceXML[1] es el estándar propuesto por el W3C para la implementación de sistemas de diálogo.

Siguiendo la definición de R.Lopez Cozar y R.Granell [I], “Los sistemas de diálogo son programas informáticos que se diseñan con la finalidad de emular a un ser humano en un diálogo oral con otra persona”.

La utilización del estándar promueve la integración de estos sistemas en los llamados 'sistemas basados en Web', así como el aseguramiento de su portabilidad y la construcción de herramientas que faciliten al usuario la definición de servicios de diálogo.

Las estructuras de estos servicios se definen a partir de entradas, que pueden consistir en gramáticas habladas o tonos DTMF[2] (Dual-tone multi-frequency) y salidas en formato de conversores texto-a-voz o grabaciones de audio.

La secuencia de interacción entre las entradas y las salidas se definen en VoiceXML[1].

VoiceXML[1] es un lenguaje de etiquetas basado en XML que permite describir servicios de voz con independencia de la aplicación en la que corran. De esta manera no es necesario conocer detalles específicos de una plataforma para entender el funcionamiento del sistema de diálogo.

El lenguaje VoiceXML[1] describe la interacción hombre-maquina a partir de los siguientes elementos:

- Salida de texto-a-voz
- Salida de audio grabado
- Reconocimiento de entrada hablada
- Reconocimiento de tonos DTMF[2]
- Grabación de entrada hablada
- Control de flujo de diálogo
- Funciones de telefonía (transferencia de llamada, desconexión, etc.).

La descripción de un servicio básico de diálogo tendría el siguiente aspecto:

```
<?xml version="1.0" encoding="UTF-8"?>
<vxml xmlns="http://www.w3.org/2001/vxml"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/vxml
   http://www.w3.org/TR/voicexml20/vxml.xsd"
   version="2.0">
  <form>
  <field name="book">
```

```
    <prompt>Would you like to listen to Murakami, Bukowski or Asimov?</prompt>
    <grammar src="book.grxml" type="application/srgs+xml"/>
  </field>
  <block>
    <submit next="http://www.library.com/book2.asp"/>
  </block>
 </form>
</vxml>
```

Este ejemplo sencillo implementaría un servicio de libros leídos donde el usuario ha de escoger entre escuchar un libro de Haruki Murakami, Charles Bukowski o Isaac Asimov. El campo field es un equivalente al campo tipo input del lenguaje HTML sólo que en este caso la entrada es hablada en lugar de tecleada. La gramática de entrada se definiría siguiendo el estándar SRGS[3] (Speech Recognition Grammar Specification). Si el sistema reconoce las palabras Murakami, Bukowski o Asimov, reproduciría un archivo de sonido con la versión en audio de un libro del autor.

La aplicación consistiría en interprete de documentos VoiceXML[1] que interactúa con una plataforma de voz, que en este caso puede ser un sistema de telefonía recogiendo las instrucciones habladas de un usuario o alternativamente la marcación por pulsos DTMF[2], como haría un sistema clásico IVR[4] (Interactive Voice Response).

2 Ventajas e Inconvenientes de VoiceXML

VoiceXML[1] es una tecnología independiente de la plataforma.

Las herramientas construidas y adaptadas para soportar el estándar comparten un formato de definición de sistemas de diálogo que permite la portabilidad y transferencia de datos entre aplicaciones heterogéneas. Además este tipo de herramientas tienden a facilitar enormemente el trabajo del desarrollador, que no necesita conocer los detalles de implementación del sistema que está construyendo.

No obstante, el seguimiento del estándar produce también inconvenientes.

VoiceXML[1] implementa funciones de telefonía, pero el conjunto de funciones es demasiado reducido y su funcionalidad muy limitada. Por ello los sistemas desarrollados con VoiceXML[1] no consiguen alcanzar la complejidad necesaria para algunos flujos de telefonía y operaciones de voz.

Como ejemplo de estas carencias se podría mencionar la imposibilidad de seleccionar el punto de una grabación donde queremos comenzar a reproducir el audio, pausarlo, rebobinarlo etc. W3C tiene previsto incorporar esta y otras características a la versión 3.0 de VoiceXML, cuyo primer borrador público está previsto para finales del año 2007.

En lo referente a las carencias en servicios de telefonía, W3C propone el estándar CCXML[5] (Call Control eXtensible Markup Language), que se puede utilizar de manera conjunta con VoiceXML[1] y que permite definir una lógica más compleja y completa del flujo telefónico, incorporando elementos necesarios como manejadores de eventos u objetos de conferencia.

Asimismo encontramos otros estándares del W3C a tener en cuenta en la construcción de una herramienta para implementaciones de sistemas de diálogo como SRGS[3] o SSML[6] (Speech Synthesis Markup Language).

Como vemos el W3C no sólo compensa los inconvenientes que pueda presentar VoiceXML[1] mediante la propuesta del uso conjunto de estándares sino que está en un proceso de permanente diálogo con la comunidad de desarrolladores para introducir ampliaciones y mejoras al estándar.

Entre las características más llamativas del futuro VoiceXML 3.0, podemos mencionar la identificación de voz, que utilizará medidas biométricas para asegurar la identidad en transacciones telefónicas y comunicaciones.

3 Voip y VoiceXML

VoiceXML[1] se utiliza en la creación de soluciones PBX (Private Branch Exchange) e IVR[4] (Interactive Voice Response) entre otras.

Aunque ya existen herramientas de desarrollo y soluciones basadas en VoiceXML[1], se podría decir que la introducción de estándares al mundo de la telefonía IP es aún parcial, pues si bien es cierto que las grandes compañías participan en la definición de los estándares y ofrecen herramientas que integran VoiceXML[1] y CCXML[5] entre otros estándares, los productos más tradicionales y los basados en código abierto, con mucho peso en la industria de las telecomunicaciones, aún no ven como algo prioritario adoptar los estándares propuestos por el W3C.

Microsoft y Sun son dos compañías permanentemente comprometidas en el desarrollo de herramientas que incorporan estándares. En el caso de VoiceXML[1], Microsoft ofrece un paquete software denominado Microsoft Speech Server[7], cuya versión 2007 aparecerá integrada en el Microsoft Office Communications Server 2007, que promueve la integración de las aplicaciones de voz con las aplicaciones basadas en Web. Por su parte Sun tiene una alianza con VoiceGenie, compañía

que dispone de su propia aplicación de telefonía denominada VoiceGenie VoiceXML Gateway[8] y cuyos planteamientos coinciden, al igual que el caso anterior, con las propuestas del W3C. También podríamos mencionar el IBM WebSphere Voice Toolkit[9] o los productos de la empresa Voxeo[10].

En el lado opuesto nos encontramos los sistemas propietarios o de distribución libre utilizados por usuarios finales y empresas de telecomunicaciones de diversa magnitud, donde la rentabilidad del negocio compite de manera más violenta con la adopción de estándares y, aunque podemos encontrar integración con módulos interpretes de VoiceXML[1], estos no se ven como una prioridad en evolución de la tecnología. Como ejemplo claro de esto podemos mencionar Asterisk, el PBX de libre distribución más utilizado del mundo, cuya integración con VoiceXML[1] aún se reduce a un módulo de libre distribución, voxy[12], cuya primera versión data de Septiembre de 2006 y algunos pocos módulos comerciales o semi-comerciales entre los que podemos mencionar VXIasterisk[13] de i6Net.

4 Conclusiones

VoiceXML[1], utilizado de manera conjunta con otros estándares, proporciona una base sólida para la definición de sistemas de voz. Las constantes revisiones y ampliaciones del estándar aseguran su continuidad y progresiva incorporación en las herramientas para la construcción de aplicaciones de voz.

El aumento de los sistemas de telefonía a través de Internet y los progresos en los campos del reconocimiento de voz y las herramientas digitales para la lectura así como la progresiva incorporación de los estándares propuestos por el W3C, señalan la gran importancia que sin duda VoiceXML[1] y el resto de estándares de voz tendrán en un futuro muy próximo.

5 Bibliografía

- [I] R. López-Cózar, R.Granell (2004). Sistemas de Diálogo Basado en VoiceXML para Proporcionar Información de Viajes en Tren. Revista: Procesamiento del Lenguaje Natural nº 33

6 Referencias Web

- [1] Voice Extensible Markup Language (VoiceXML) Version 2.0 - W3C Recommendation 16 March 2004
  http://www.w3.org/TR/voicexml20/
- [2] Dual-tone multi-frequency (DTMF)
  Wikipedia
  http://en.wikipedia.org/wiki/DTMF
- [3] Speech Recognition Grammar Specification (SRGS)
  Version 1.0 W3C Recommendation 16 March 2004
  http://www.w3.org/TR/speech-grammar/
- [4] Interactive Voice Response (IVR)
  Wikipedia
  http://en.wikipedia.org/wiki/Interactive_voice_response
- [5] Call Control eXtensible Markup Language (CCXML)
  Version 1.0 - W3C Working Draft 30 April 2004
  http://www.w3.org/TR/2004/WD-ccxml-20040430/
- [6] Speech Synthesis Markup Language (SSML)
  Version 1.0 - W3C Recommendation 7 September 2004
  http://www.w3.org/TR/speech-synthesis/
- [7] Microsoft Speech Server
  http://www.microsoft.com/speech/default.mspx
- [8] VoiceGenie VoiceXML Gateway
  http://developer.voicegenie.com/platforminfo_VoiceGenie.php?page=vgtelserver
- [9] IBM WebSphere Voice Toolkit http://www-306.ibm.com/software/pervasive/voice_toolkit/"
- [10] Voxeo http://www.voxeo.com
- [11] Asterisk http://www.asterisk.org/
- [12] Voxy http://sourceforge.net/projects/voxy
- [13] VXIasterisk
  http://products.i6net.com/index.php?tg=entry&idx=view&article=37