



LOS PORTALES DE VOZ: INTERNET MAS NATURAL

Resumen

Cada vez está más cerca el día en que Internet conversará a viva voz con la gente. Varias empresas ya anunciaron nuevos voice browsers (navegadores por voz) que permitirán acceder a los contenidos de la Web sin necesidad de una computadora, Por ejemplo utilizando un teléfono común o desde un celular. Además, este tipo de navegadores facilitará el acceso a Internet a personas con dificultades visuales o motoras, y a aquellas que no poseen un computador.

Palabras Claves: Portales de voz, VoiceXML, VoIP, TTS.

1. Introducción

El acceso a Internet por medio de la voz debería ser el método más natural e inherente para un ser humano. En realidad, no habría nada más sencillo que navegar por Internet, interactuando mediante órdenes vocales, de manera similar a una conversación normal.

El concepto de Portal de Voz, actualmente también denominado Internet Universal, da un paso más allá de la pura navegación vocal y pretende relacionar el mundo Internet con el mundo de la voz, a través de un conjunto de servicios que amplíen, complementen y den un valor agregado al acceso vocal a Internet propiamente dicho.

Gracias a los Portales de Voz, se establece la convergencia del mundo Internet con el mundo de la telefonía vocal más elemental, permitiendo el acceso no sólo a los clientes del servicio telefónico fijo, sino también a los del servicio móvil.

En la Figura 1 se muestra la evolución histórica del acceso a Internet: inicialmente, sólo se tenía acceso a Internet a través de páginas en formato HTML, utilizando un web browser, con la aparición de la telefonía móvil celular se logró tener otro medio de acceso a Internet empleando el protocolo WAP y el formato WML, y por último, la innovación con los portales de voz a través de un teléfono móvil o fijo, o inclusive desde el mismo PC, a través del formato VoiceXML.

Este artículo pretende dar una visión general de lo que es un Portal de Voz, su estructura, el lenguaje en el que está soportado: VoiceXML, y los avances en este reciente campo de investigación.

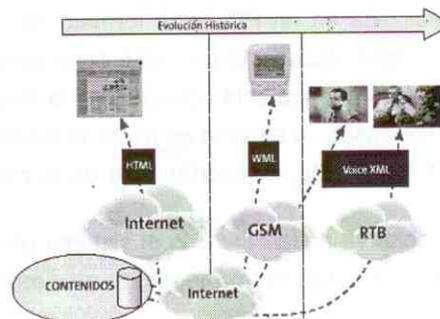


Fig. 1. Evolución del acceso a Internet

•DEFINICIÓN DE UN PORTAL DE VOZ

Los portales de voz son sitios web diseñados generalmente, bajo el lenguaje VoiceXML, que permiten navegar mediante el reconocimiento de lenguaje natural, por un variado menú de contenidos, consultar información y conectar con comercios, empresas o números de teléfono solicitados, con solo mencionar su nombre. La forma de acceso es una simple marcación telefónica y la infraestructura

necesaria, un teléfono fijo o móvil.

3. ESTRUCTURA DE UN PORTAL DE VOZ

Se requiere una plataforma que suministre servicios de usuario y de red, que solicitan recursos especializados (voz, fax, datos, etc.) y/o recursos de comunicaciones (operadoras, protocolos de señalización, protocolos de comunicación, etc.). En un entorno de operación por RTB (Red Telefónica Básica) o por VoIP (Voz sobre IP), el acceso al servicio del Portal de Voz se realiza mediante una llamada de teléfono. La plataforma funciona como elemento intermediario, estableciendo la conexión entre el usuario que realiza la llamada y el servicio. En un primer escenario, se realiza una llamada tradicional por la RTB, conducida hacia la plataforma de acceso vocal a Internet donde accede al servicio de Portal de Voz. En un segundo escenario, el primer tramo de la llamada se realiza por la RTB, pero, tras ser transformada por el *gateway* en voz sobre IP, se transmite hacia la plataforma de acceso vocal a Internet (conectada en este caso directamente a la red IP). Finalmente, una llamada que es realizada en un PC, tiene formato de VoIP desde su origen, accediendo en este caso también hacia la plataforma que está conectada a la Red IP. En todos los casos, el servicio es ajeno al medio de acceso de la llamada, comportándose de la misma forma.

Como se observa en la Figura 2, el sistema para la prestación del servicio de voz se realiza mediante los bloques funcionales siguientes:

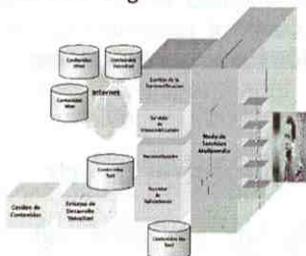


Fig. 2. Sistema para la prestación del Servicio de Voz

•**Los usuarios:** Realizan las llamadas telefónicas al servicio, solicitando unos determinados contenidos en su navegación por el Portal de Voz.

•**El Nodo de Servicios Multimedia:** Esta plataforma es capaz de prestar múltiples servicios. En concreto, se encargará de buscar los contenidos VoiceXML solicitados durante la navegación vocal para presentárselos al usuario. Soportando los protocolos más habituales de señalización (PUSI, PARI, DSS1, SMS, etc.) del mundo Internet (SMTP, POP3, etc.), la conexión a múltiples redes (IP, Móviles, RTB, etc.), así como el acceso por Voz sobre IP mediante el protocolo H.323 y la conexión a bases de datos. Para la prestación del servicio de acceso vocal a Internet, de acuerdo a la clase de contenidos (propios o de Internet) y su formato, la plataforma se apoyará para interpretarlos, en algunos de los bloques funcionales que se describen a continuación.

•**El Servidor de Aplicaciones:** Proporciona acceso a contenidos de cualquier origen y de ellos extrae los datos puntuales que necesita el servicio que esté ofreciendo la plataforma. En este caso, los datos proporcionados a la plataforma no están en el VoiceXML debido a que es la propia plataforma la que, realiza la conversión a voz de estos datos puntuales (por ejemplo la lectura de *e-mail*). El servidor de aplicaciones podrá estar distribuido físicamente en tantas máquinas como sea necesario para la prestación eficaz del servicio.

•**El Servidor de Transcodificación:** Permite convertir a formato VoiceXML contenidos de cualquier origen (basados en XML).

•**El módulo de Gestión de la Transcodificación:** Este módulo permite asociar un conjunto de patrones de transformación aplicables a una URL concreta, para facilitar y automatizar la manera en que se realiza la conversión de cada uno de los formatos. Esta gestión se realiza de una forma gráfica y amigable, lo cual facilita la definición de dichos patrones.

•**El módulo de Personalización:** Con este módulo se permite la generación dinámica de páginas VoiceXML. Dentro de la plataforma de acceso vocal a Internet, la reproducción de las páginas VoiceXML se realiza mediante un componente de la plataforma que se denomina *Browser Vocal*.

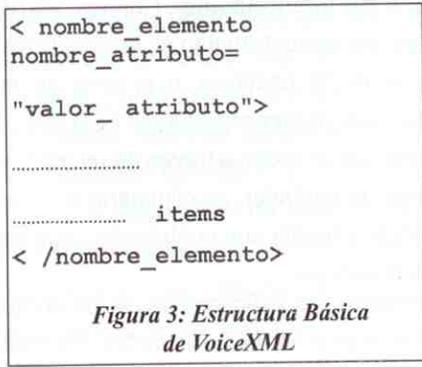


4. EL LENGUAJE VOICEXML

El lenguaje VoiceXML, forma parte de la familia XML (Extensible Markup Language), y está enmarcada hacia la creación de aplicaciones de reconocimiento automático del habla y respuesta interactiva por voz.

VoiceXML nació con el proyecto PhoneWeb en los laboratorios Bell de AT&T. Después de la división de esa compañía en AT&T por un lado, y Lucent Technologies, por otro, el proyecto continuó en Lucent, pero con el nombre de TelePortal. Otro aporte a VoiceXML vino de parte de la empresa Motorola, la cual en el año 1998 crea el VoxML (Voice Markup Language), lo cual sirvió como base para que años más tarde en unión con AT&T, Lucent, IBM y W3C (World Wide Web Consortium) crearan el VoiceXML Forum, un sitio que sigue paso a paso la evolución de la tecnología de reconocimiento de la voz en Internet. Estas empresas y otras se han unido a través del tiempo tomando este lenguaje como estándar, para la creación de navegadores y servicios online a los que se puede acceder por medio de la voz y desde un teléfono. Actualmente se tiene la versión 2.0, de las especificaciones de este lenguaje.

Al igual que su 'padre' XML, está basado en tags, los cuales, encierran instrucciones llamadas item, como lo muestra la figura 3. Un tag, puede tener atributos; cada atributo consta de un nombre y un valor, separados por el signo igual (=); además, el valor debe estar encerrado en comillas.



Una aplicación VoiceXML, consiste en uno o más archivos de texto, llamados documentos, los cuales

son denotados por medio de la extensión *.vxml*

Un ejemplo de VoiceXML, se presenta a continuación en la Figura 4. Este documento proporciona respuesta a la pregunta: "Por favor, escoja Noticias, Clima o Deportes", seguidamente este documento acepta la selección del usuario y pasa el control a otro llamado "seleccionar.jsp", el cual es un script del lado del servidor (server-side), que seguramente proveerá el servicio que el usuario ha elegido.

Ejemplo de atributos son los de las líneas 3 y 5 que muestran los atributos versión y nombre, respectivamente.

<p>Li</p> <p>1</p> <p>2</p> <p>3</p> <p>4</p> <p>5</p> <p>6</p> <p>7</p> <p>8</p> <p>9</p> <p>10</p> <p>11</p> <p>12</p> <p>13</p> <p>14</p> <p>15</p> <p>16</p> <p>17</p>	<p>Código VoiceXML</p> <pre> <?xml version="1.0"?> <!DOCTYPE vxml PUBLIC "-//BeVocal Inc//VoiceXML 2.0//EN" "http://cafe.bevocal.com/libraries/ dtd/vxml2-0-bevocal.dtd"> <vxml version="2.0"> <form> <field name="seleccion"> <prompt> Por favor escoja Noticias, Clima o Deportes </prompt> <grammar> [Noticias Clima Deportes] </grammar> </field> <block> <submit next="seleccionar.jsp"/> </block> </form> </vxml> </pre>	<p>TTS</p>
--	---	-------------------

Figura 4. Un ejemplo de VoiceXML

Como se puede ver, los tags vienen en pares que se distinguen por el símbolo /, por ejemplo las líneas 4 y 16. Estas parejas son llamadas Contenedores.

Existen otros tipos de tags, los cuales no tienen pareja y son denominados stand-alone o tags vacíos. Un



ejemplo de estos se encuentra en la línea 14 y se distinguen porque justo antes del carácter de cerrado (>) tiene un slash (/).

Las líneas 1, 2 y 3 son requeridas al comienzo de todo documento VoiceXML. Estas describen las versiones de las especificaciones que se utilizarán y la ruta del archivo DTD (Document Type Definition) que se va a utilizar, el cual contiene la definición y la manera de interpretar todos los tags del documento.

En VoiceXML se utilizan los Tags <form>, <field> y <submit> para crear una caja de selección y un botón; Similar a HTML donde se utiliza <form>, <select> e <input>.

Un Formulario, generalmente, contiene campos de información que el usuario puede seleccionar. En el ejemplo, el formulario tiene un campo llamado selección (línea 5), donde se almacena el tipo de información (para este caso) que el usuario desea consultar. El tag <prompt> (línea 6) contiene un texto que está a manera de etiqueta. El intérprete de VoiceXML usará su sistema de conversión TTS (Text-To-Speech) para leer este texto y enviarlo al usuario en forma de audio. En la línea 9, está el tag <grammar> que contiene las tres posibles respuestas que la plataforma aceptará, (línea 10). Seguidamente se realiza el proceso contrario; se recibe un archivo de audio el cual es convertido a una de estas selecciones (suponiendo que se haga de forma correcta) y se pasa el control a otro archivo encargado de presentar los contenidos (línea 14).

5. AVANCES EN EL CAMPO

El deseo de dominar el mercado de la voz en Internet ha desatado una competida carrera entre las grandes empresas. Compañías como Intel, IBM, Yahoo y America Online ya están tomando posiciones para desarrollar sus propios portales de voz.

El despegue de los portales de voz se produjo en los Estados Unidos. Las empresas que se iniciaron en este campo, entre otras, fueron: Tellme, TelSurf, además de BeVocal, Audiopoint y Speaklink, que también han lanzado sus propios portales de voz AT&T invirtió 60 millones de dólares en Tellme

Networks para convertirse en el principal proveedor de esta empresa en materia de servicios de comunicaciones. Lycos lanzó a finales del 2000 un portal con servicios y contenidos basados en la voz, mediante una alianza con Quack.com, que pone la infraestructura y el soporte del sitio; la empresa española People Communications tiene en marcha Ydilo, un portal de voz para España y Latinoamérica.

Varias empresas han anunciado ya nuevos navegadores de voz que permitirán acceder a los contenidos de la red sin necesidad de un PC, por ejemplo hablando por, un teléfono común o desde un celular. Entre estos se encuentran el Nuance Voyager y el SpeechSite, dos programas que sirven para crear sitios web capaces de interpretar el lenguaje natural y cotidiano que utilizan las personas y permiten el acceso a Internet a través de un teléfono. Otro navegador de voz destacado es el Conversa Web, aunque a diferencia de los anteriores, este software está diseñado para navegar con comandos de voz desde un computador.

Con respecto al establecimiento de estándares para esta nueva tecnología, la empresa Nuance, que fabrica el navegador de voz Voyage, encabeza un consorcio formado por más de veinticinco compañías cuyo objetivo es fijar un estándar de seguridad para los nodos de comercio electrónico por voz. Algunos miembros del consorcio, llamado V-commerce, quieren que sus tecnologías de voz sean capaces de reconocer a sus interlocutores. Nuance, VeriVoice y otras firmas del sector trabajan en la identificación de aspectos como la cadencia o el tono de la voz, entendidos como marcas que sirven para identificar a los usuarios que acceden a través de un teléfono. De conseguirse un estándar, se eliminaría la necesidad de memorizar y repetir una contraseña cada vez que se accede al servicio.

Por el momento, las aplicaciones de los portales de voz son escasas y no muy relevantes. Permiten por ejemplo leer el correo, obtener información sobre deportes o la bolsa o consultar los servicios de las líneas aéreas. Una vez que la seguridad de las transacciones orales esté plenamente garantizada,



estos nodos pretenden adentrarse en el mercado del comercio electrónico. Ventajas hay para ello: no hay que teclear ni observar pantallas, sólo hablar; se eliminan las barreras físicas para la gente que no puede leer o escribir, con lo que se puede fomentar el acceso a la red en los países menos desarrollados y para las personas discapacitadas; y ni siquiera hace falta contar con un PC: basta con realizar una llamada telefónica.

6. VIABILIDAD

Varias compañías se preparan para tener una porción del mercado de los servicios de voz de Internet, gracias al estándar del formato VoiceXML.

Según la consultora Gartner Group, más de un 30 por ciento de esos proyectos estarán en pleno funcionamiento en el año 2003. La consultora también predice que el crecimiento mayor se dará en los portales de voz que brinden servicio de correo de voz, noticias, pronóstico del tiempo y cotizaciones de la bolsa, por ejemplo. Para esto se deben alcanzar los niveles de seguridad que en estos momentos se encuentran las consultas y transacciones que se hacen por un PC común y corriente, si se tiene en cuenta que los servicios anteriormente dichos se harán por voz.

En nuestro país no es aún muy claro el conocimiento que se tiene acerca de la aparición de los portales de voz. Las empresas telefónicas que compiten en el mercado los ven como una alternativa que no tiene mucha fuerza, por lo que se encuentran apenas en etapa de estudio, además, parece ser que no existe ninguna regulación normativa al respecto por parte de la Comisión Reguladora de Telecomunicaciones. La única manera de tener un servicio de portal de voz es contratándolo con alguna de las empresas internacionales que ya son muy fuertes en este mercado.

•REFERENCIAS

http://www.ts.es/soluciones/sol_crm/Portal%20de%20Voz.PDF
F
<http://www.wmlclub.com/articulos/voz.htm>
<http://www.w3.org/TR/2000/NOTE-voicexml-20000505/#toc>
<http://cafe.bevocal.com/docs/tutorial/index.html>
http://www.momentum-tech.com/voice_wp.pdf

AUTORES

Lola Xiomara Bautista
Marlene Guerrero
Nidia Johanna rey
Hernán Darío Rojas
José Luis Barbosa

