



Universidad Carlos III

I.T. Informática de Gestión

Proyecto Fin de Carrera

Aplicación Práctica del estándar VoiceXML para el desarrollo de un Sistema de Diálogo Oral

Autor: Oscar Flores González

Tutor: David Griol Barres

Co-Director: Javier Ignacio Carbó Rubiera

Índice

- ➔ Objetivos
- ➔ Sistemas de Diálogo
- ➔ Sistemas de Diálogo Orales
- ➔ El estándar VoiceXML
- ➔ Aplicación Realizada
- ➔ Demostración Práctica y Estadísticas de Uso
- ➔ Conclusiones
- ➔ Ruegos y preguntas

- ➔ Objetivos
- ➔ Sistemas de Diálogo
- ➔ Sistemas de Diálogo Orales
- ➔ El estándar VoiceXML
- ➔ Aplicación Realizada
- ➔ Demostración Práctica y Estadísticas de Uso
- ➔ Conclusiones
- ➔ Ruegos y preguntas

Objetivos

- Aplicar el estándar VoiceXML a un Sistema de Diálogo Oral.
- ¿Qué es un Sistema de Diálogo?
- VoiceXML: estándar para la creación de Sistemas de Diálogo Orales.
- Aplicación práctica: orientada al mundo de la simulación aérea, proporcionará información meteorológica y de las pistas de un aeropuerto.

- ➔ **Objetivos**
- ➔ **Sistemas de Diálogo**
- ➔ Sistemas de Diálogo Orales
- ➔ El estándar VoiceXML
- ➔ Aplicación Realizada
- ➔ Demostración Práctica y Estadísticas de Uso
- ➔ Conclusiones
- ➔ Ruegos y preguntas

Sistemas de Diálogo

- ➔ Permiten a las personas interactuar con sistemas informáticos.
- ➔ Calsificación según:
 - ➔ Modalidad.
 - ➔ Dispositivo.
 - ➔ Estilo.
 - ➔ Iniciativa de diálogo.
 - ➔ Función.

- ➔ Objetivos
- ➔ Sistemas de Diálogo
- ➔ **Sistemas de Diálogo Orales**
- ➔ El estándar VoiceXML
- ➔ Aplicación Realizada
- ➔ Demostración Práctica y Estadísticas de Uso
- ➔ Conclusiones
- ➔ Ruegos y preguntas

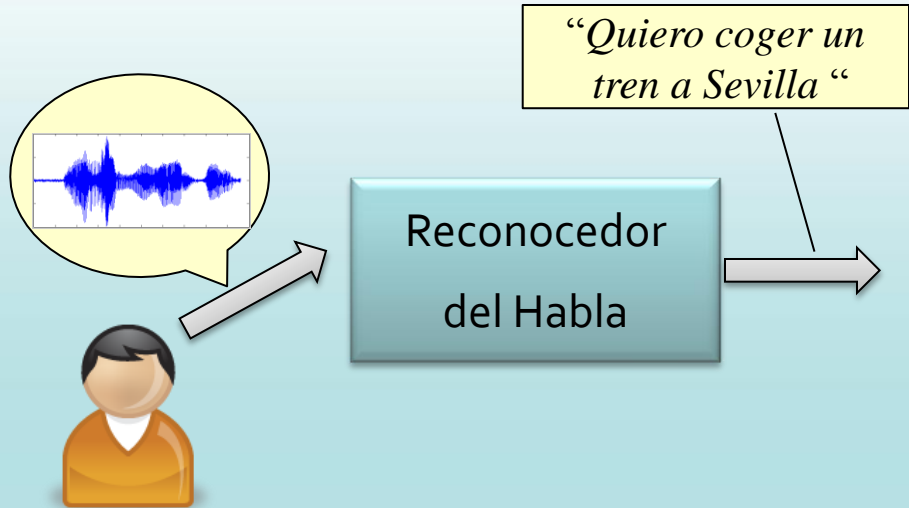
Sistemas de Diálogo Orales

- ➔ Voz: es la forma natural de comunicación entre los seres humanos.
- ➔ Los sistemas de diálogo orales permiten la interacción hombre-máquina a través de la voz.
- ➔ Componentes:
 - ➔ Reconocedor del Habla.
 - ➔ Módulo de Comprensión del Habla.
 - ➔ Gestor del diálogo.
 - ➔ Sintetizador de Voz.
 - ➔ Sistema Gestor de Base de Datos.

Sistemas de Diálogo Orales

Reconocedor del Habla

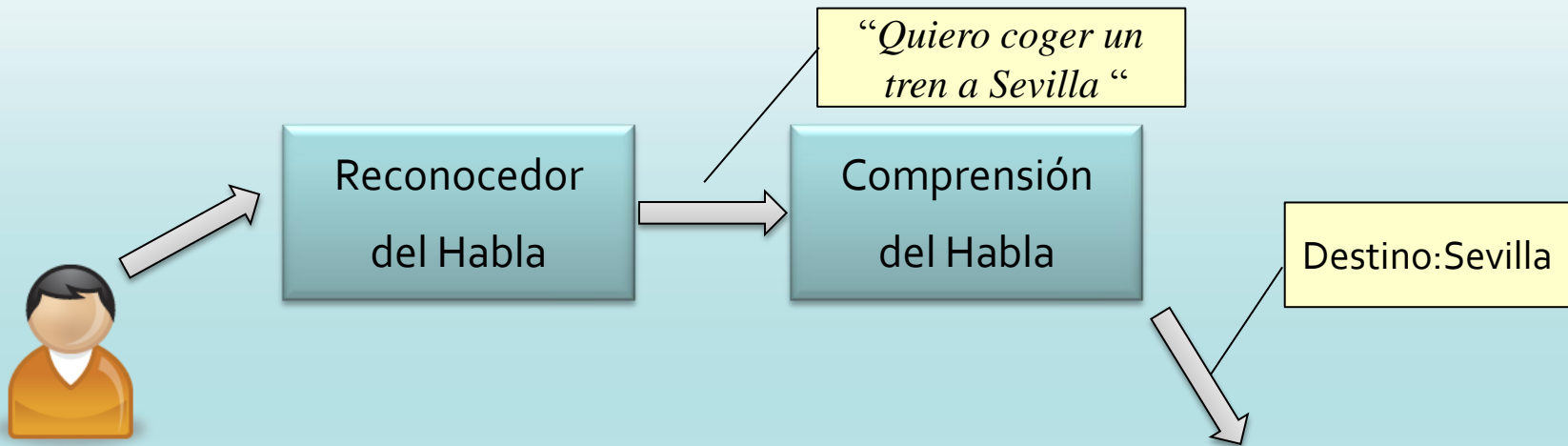
- ➔ Recibe y procesa la voz recibida del usuario



Sistemas de Diálogo Orales

Módulo de Comprensión del Habla

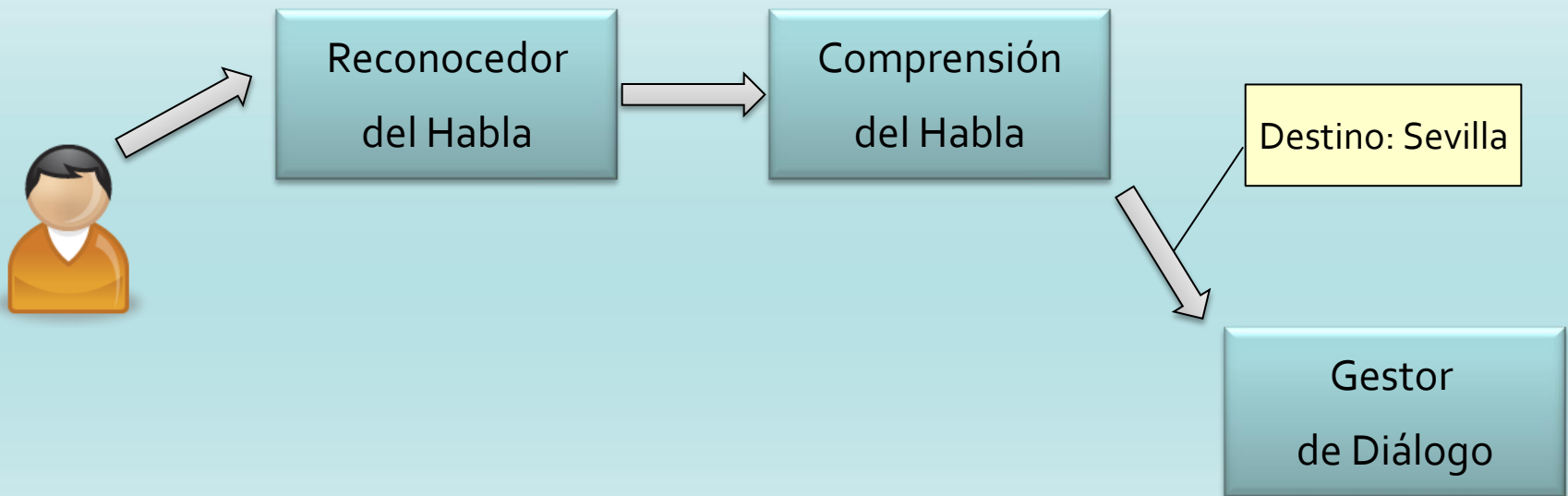
- Procesa las palabras generadas por el ASR y obtiene su representación semántica



Sistemas de Diálogo Orales

Gestor de Diálogo

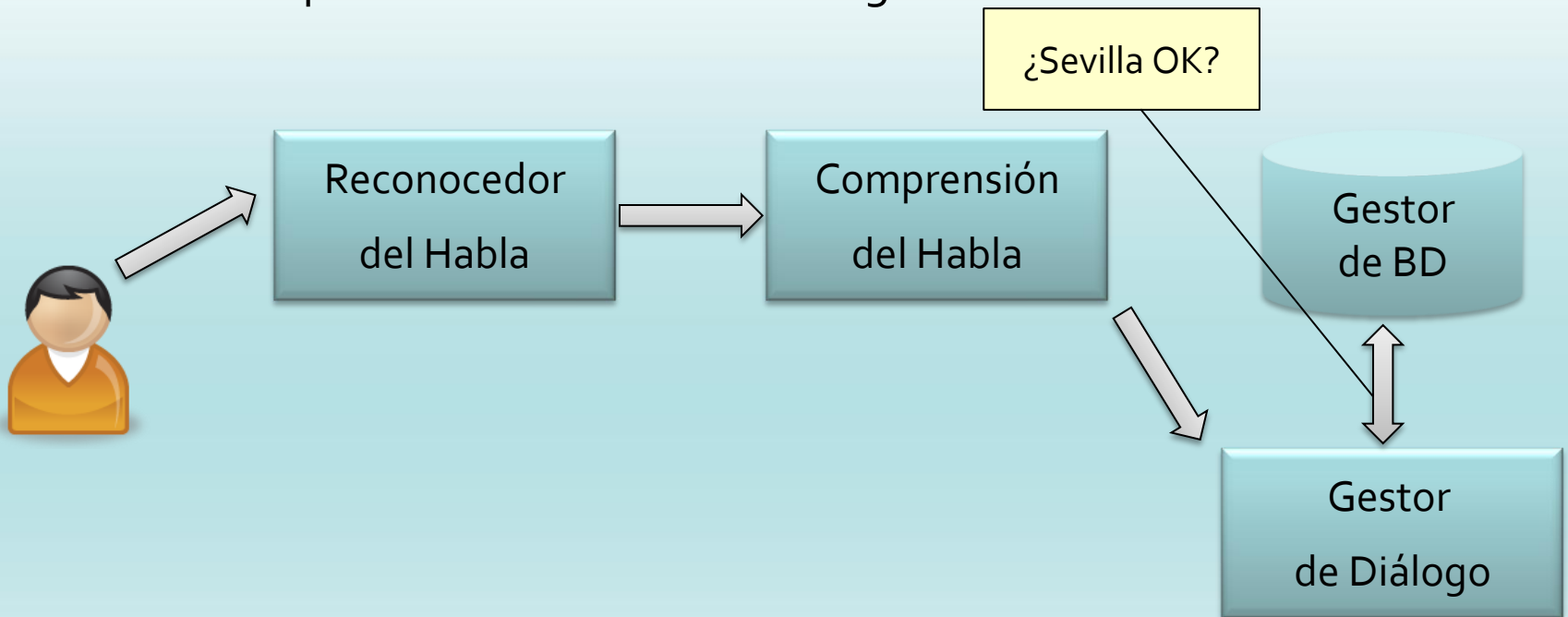
- Recibe la representación semántica obtenida por el Módulo de Comprensión.
- Es el elemento más importante.



Sistemas de Diálogo Orales

Sistema Gestor de Base de Datos

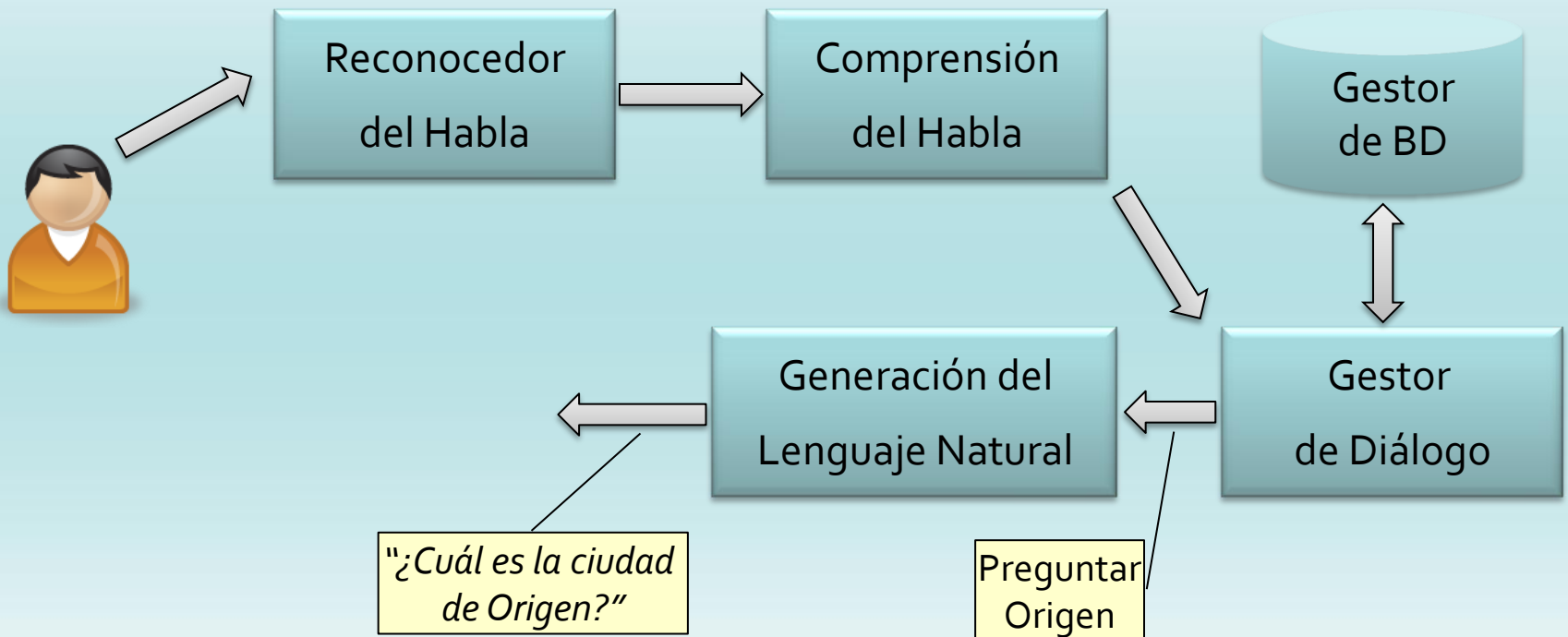
- ➔ Accede a la información almacenada en la Base de Datos a petición del Gestor de Diálogo



Sistemas de Diálogo Orales

Módulo de Generación del Lenguaje Natural

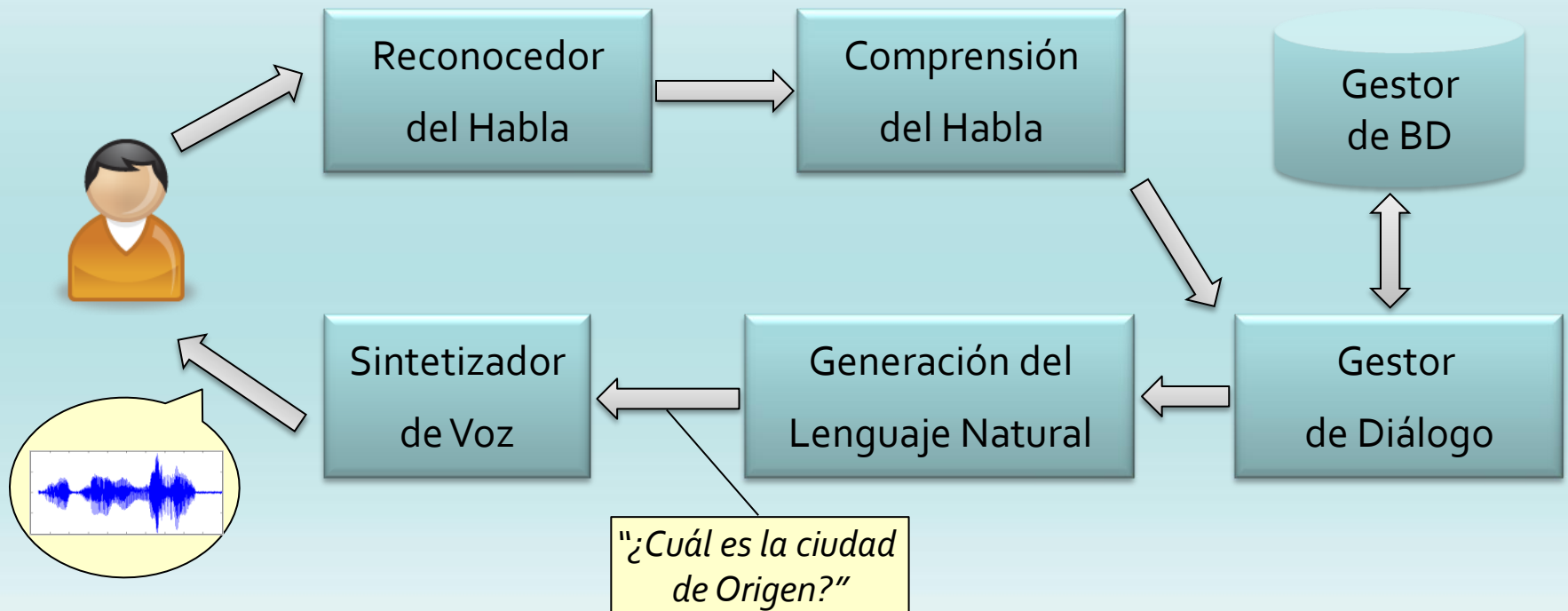
- Genera una frase en lenguaje natural según decida el Gestor de Diálogo



Sistemas de Diálogo Orales

Sintetizador de Voz

- ➔ Transforma las palabras o frases de la respuesta al usuario en voz sintetizada



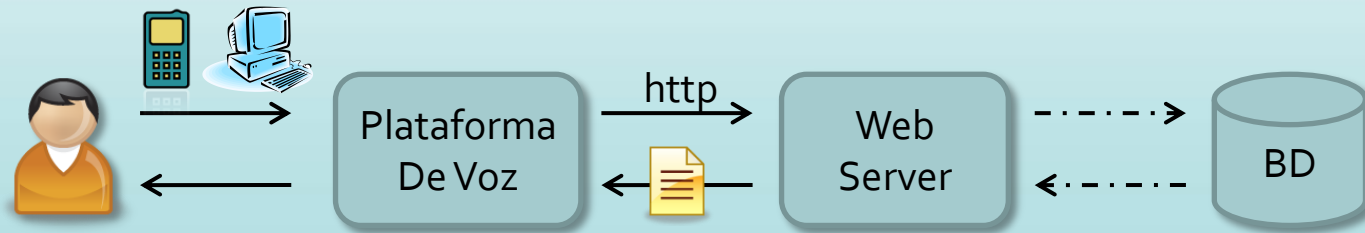
- ➔ Objetivos
- ➔ Sistemas de Diálogo
- ➔ Sistemas de Diálogo Orales
- ➔ El estándar VoiceXML
- ➔ Aplicación Realizada
- ➔ Demostración Práctica y Estadísticas de Uso
- ➔ Conclusiones
- ➔ Ruegos y preguntas

El estándar VoiceXML

- Es el estándar definido por el W3C específico para la interacción hombre-máquina en diálogos de voz.
- Publicado por IBM, Lucent, Motorola y AT&T en 2000.
- Análogo a HTML.
- Permite generar documentos con los que se interactúa a través de la voz.
- Ofrece:
 - Salida de voz sintetizada.
 - Salida de ficheros de audio.
 - Reconocimiento de entradas por voz y DTMF.
 - Grabación de diálogos.
 - Control del flujo del diálogo.
 - Funciones de telefonía.

El estándar VoiceXML

➤ Arquitectura típica:



El estándar VoiceXML

➔ Ejemplo:

```
<?xml version="1.0"?>
```

```
<vxml version="2.0">
```

```
<form id="form_nombre">
```

¿Qué queremos saber?

```
<field name="nombre">
```

Pregunta

```
<prompt> ¿Cuál es su nombre? </prompt>
```

Posibles Respuestas

```
<grammar> Juan | María | Pedro </grammar>
```

Sin Respuesta

```
<noinput> Por favor, diga su nombre. </noinput>
```

Respuesta Incorrecta

```
<nomatch> No le he entendido. </nomatch>
```

Respuesta Correcta

```
<filled>
```

```
Gracias, <value expr="nombre" />
```

¿Qué hago ahora?

```
<submit next="siguiente_doc.vxml" />
```

```
</filled>
```

```
</field>
```

```
</form>
```

```
</vxml>
```

- ➔ **Objetivos**
- ➔ Sistemas de Diálogo
- ➔ Sistemas de Diálogo Orales
- ➔ El estándar VoiceXML
- ➔ **Aplicación Realizada**
- ➔ Demostración Práctica y Estadísticas de Uso
- ➔ Conclusiones
- ➔ Ruegos y preguntas

Aplicación Realizada

- Objetivo del Proyecto: aplicar el estándar VoiceXML en un sistema de diálogo oral.
- Aplicación a desarrollar: servicio de ayuda a pilotos virtuales y aficionados a la simulación aérea.
- Dos servicios de voz en uno:
 - Información meteorológica de un aeropuerto.
 - Información de pistas de un aeropuerto.
- El usuario elegirá el aeropuerto diciendo las 4 letras (del alfabeto fonético aeronáutico) de su código ICAO (International Civil Aviation Organization).
 - Ej: Madrid Barajas ➡ LEMD (Lima – Eco – Mike – Delta)

Aplicación Realizada

➤ Información meteorológica:

➤ Necesaria para realización de un vuelo (virtual o real).

➤ METAR: METeorological Aerodrome Report, información meteorológica codificada de un aeropuerto.

➤ Ejemplo:

```
LEMD 051230Z 01014KT 340V060 9999 FEW050 14/M01 Q1012 NOSIG
```

Aplicación Realizada

➡ Significado:

```
LEMD 051230Z 01014KT 340V060 9999 FEW050 14/M01 Q1012 NOSIG
```



Código del Aeropuerto: LEMD: Madrid-Barajas

Aplicación Realizada

➡ Significado:

LEMD 051230Z 01014KT 340V060 9999 FEW050 14/M01 Q1012 NOSIG



Información obtenida el día 5 a las 12:30 h Zulú

Aplicación Realizada

➡ Significado:

LEMD 051230Z 01014KT 340V060 9999 FEW050 14/M01 Q1012 NOSIG



Viento de rumbo 010 grados. Velocidad 14 nudos. Viento con dirección variable entre 340 y 060 grados.

Aplicación Realizada

➡ Significado:

LEMD 051230Z 01014KT 340V060 9999 FEW050 14/M01 Q1012 NOSIG



Visibilidad: 10 kilómetros o más.

Aplicación Realizada

➡ Significado:

```
LEMD 051230Z 01014KT 340V060 9999 FEW050 14/M01 Q1012 NOSIG
```



Nubes: Escasas a 5000 pies sobre el nivel del aeródromo.

Aplicación Realizada

➡ Significado:

```
LEMD 051230Z 01014KT 340V060 9999 FEW050 14/M01 Q1012 NOSIG
```



*Temperatura: 14 grados Celsius.
Punto de rocío: 1 grado Celsius bajo cero*

Aplicación Realizada

➡ Significado:

LEMD 051230Z 01014KT 340V060 9999 FEW050 14/M01 **Q1012** NOSIG



QNH: 1012 Hectopascals.

Aplicación Realizada

➡ Significado:

LEMD 051230Z 01014KT 340V060 9999 FEW050 14/M01 Q1012 **NOSIG**

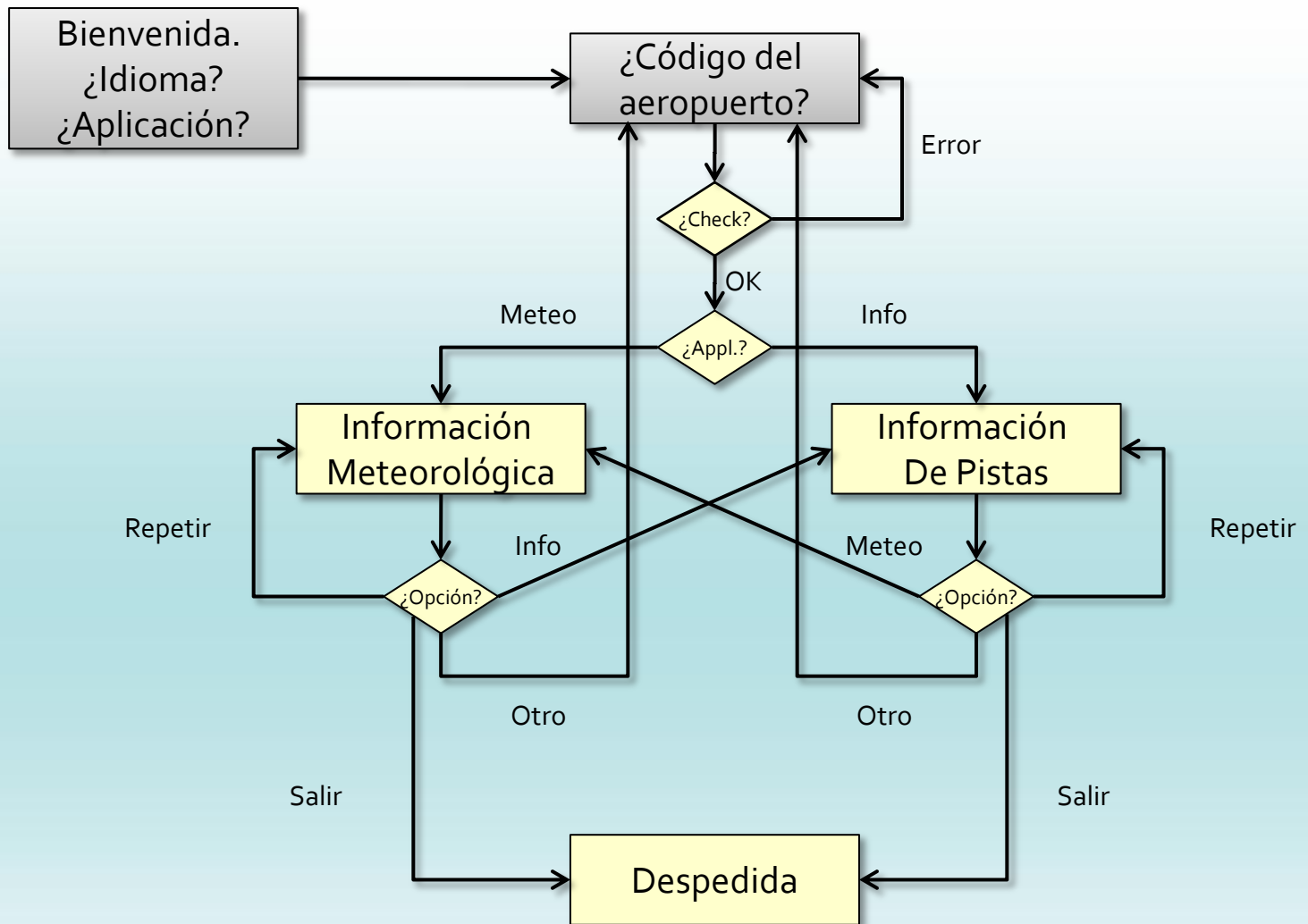


Sin cambios significativos.

Aplicación Realizada

- Información de pistas:
 - Útil para saber procedimiento en el aterrizaje.
- El servicio desarrollado:
 - Preguntará un código de aeropuerto al usuario.
 - Obtendrá la información del aeropuerto en la base de datos.
 - Devolverá la información al usuario en forma de voz.

Aplicación Realizada



Aplicación Realizada

- Entorno de desarrollo:
 - VoiceXML: diálogos estáticos.
 - Java: Servlets y clases para generación de diálogos VoiceXML de forma dinámica.
 - Apache Tomcat: servidor web para alojamiento de Servlets de la aplicación.
 - MySQL: Base de Datos relacional con información de aeropuertos (accedida desde los servlets). 2 tablas:
 - Airports.
 - Runways.
- Voxeo: Plataforma de voz para administrar y configurar la aplicación (acceso mediante Skype).
- Otros: Eclipse, Ant, Subversion.

Aplicación Realizada

- VoiceXML estáticos:
 - Documentos que no necesitan ser generados de forma dinámica.
 - `initVoiceService.vxml`
 - Punto de entrada a la aplicación.
 - Permite elegir idioma y subaplicación
 - `enterAirportCode.vxml`
 - Permite al usuario introducir el código del aeropuerto sobre el que se devolverá la información solicitada (meteorología o información general).

Aplicación Realizada

- Gramática `phonetic_es.grxml`:
 - Principal gramática de la aplicación.
 - Necesaria para aceptar códigos de aeropuerto.
 - Proporciona flexibilidad.
 - Consta de 2 reglas principales:
 - Rule *icao_code*
 - Rule *letter*

Aplicación Realizada

➔ Regla icao_code:

➔ Regla raíz (root) de la gramática.

```
<rule id="icao_code" scope="public">
  <tag>out.finalcode="";</tag>
  <item repeat="4">
    <ruleref uri="#letter"/>
    <tag>out.finalcode = out.finalcode + rules.letter.id_letter</tag>
  </item>
  <tag>out.option = out.finalcode;</tag>
</rule>
```

➔ Regla letter:

```
<rule id="letter">
  <one-of>
    <item>alfa<tag>out.id_letter = "A";</tag></item>
    <item>bravo<tag>out.id_letter = "B";</tag></item>
    <item>charli<tag>out.id_letter = "C";</tag></item>
    ...
    <item>victor<tag>out.id_letter = "V";</tag></item>
    <item>uiski<tag>out.id_letter = "W";</tag></item>
    <item>exrey<tag>out.id_letter = "X";</tag></item>
    <item>yanqui<tag>out.id_letter = "Y";</tag></item>
    <item>zulu<tag>out.id_letter = "Z";</tag></item>
  </one-of>
</rule>
```

Aplicación Realizada

- Ventaja principal: permite reconocer el código de cualquier aeropuerto de forma sencilla.
- En lugar de una sola regla con un ítem por cada código posible...

```
<item>AAAA</item>  
<item>AAAB</item>  
<item>AAAC</item>  
...  
<item>ZZZY</item>  
<item>ZZZZ</item>
```

- ...que implicaría un total de 23751 ítems:

(Combinaciones con repetición de 26 elementos tomados de 4 en 4)

$$\left(\frac{n+m-1!}{m! (n-1)!} \right) \Rightarrow \left(\frac{29!}{4! \cdot 25!} \right) = 23751$$

- ...sólo son necesarias dos reglas y 26 ítems (uno por letra).

Aplicación Realizada

- VoiceXML generados dinámicamente:
 - Generados por los Servlets, lo que permite:
 - Generar documentos Vxml en el idioma que corresponda, documentos Vxml de error.
 - Acceder a fuentes de información externas e internas.
 - Generar el mensaje respuesta al usuario en función de la entrada recibida.

Aplicación Realizada

➔ CheckAirportCode:

- ➔ Busca el código del aeropuerto en BD.
- ➔ Genera un documento VoiceXML en el que el usuario confirma que es el aeropuerto que desea.

Aplicación Realizada

➤ GetMetar:

- Obtiene de un servidor externo la información meteorológica (sin codificar) del aeropuerto elegido.
- Transforma la información en texto entendible por el usuario.
- Ofrece al usuario:
 - Volver a escuchar la información.
 - Buscar información meteorológica de otro aeropuerto.
 - Buscar información de pistas de ese mismo aeropuerto.
 - Salir.

Aplicación Realizada

➤ GetAirportInfo:

- Obtiene de la BD información de las pistas referente a un aeropuerto.
- Transforma la información en texto entendible por el usuario.
- Ofrece al usuario:
 - Volver a escuchar la información.
 - Buscar información de pistas de otro aeropuerto.
 - Buscar información meteorológica de ese mismo aeropuerto.
 - Salir.

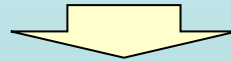
Aplicación Realizada

➤ Otras Clases Java:

➤ Decoder

- Transforma un informe METAR en texto inteligible.
- Basado en expresiones regulares

```
^([\d-]{3}|VRB)([\d-]{2,3})(G[\d-]{2,3})?(KT|MPS|KMH)$
```



29011KT: Viento de 290. Velocidad 11 nudos

VRB03KT: Viento con dirección variable. Velocidad 3 nudos.

23004MPS: Viento de 230. Velocidad 4 metros/seg.

24035G55KT: Viento de 240. Velocidad 35 nudos con rachas de 55 nudos.

...



- ➔ Objetivos
- ➔ Sistemas de Diálogo
- ➔ Sistemas de Diálogo Orales
- ➔ El estándar VoiceXML
- ➔ Aplicación Realizada
- ➔ Demostración práctica y Estadísticas de Uso
- ➔ Conclusiones
- ➔ Ruegos y preguntas

Demostración Práctica

(Plataforma VOXEO)



+990009369991433214

-  Información Meteorológica (Castellano, aeropuerto KJFK, New York)
-  Información De Pistas (Inglés, aeropuerto LEST, Santiago de Compostela)

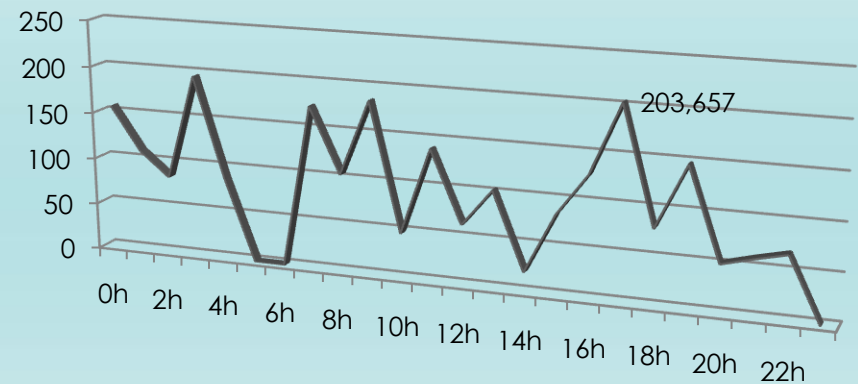
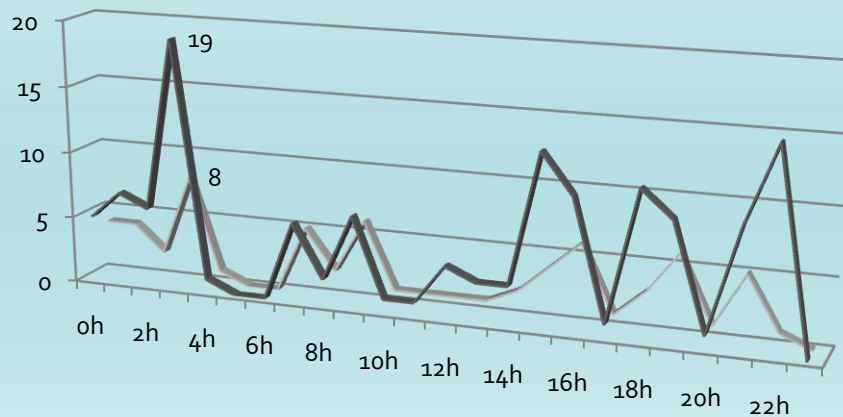
Estadísticas de Uso

- Ejemplo de estadísticas de uso:
 - Información obtenida después de analizar manualmente logs proporcionados por Voxeo.
 - Datos desglosados por hora referentes a:
 - Duración media de llamadas.
 - Llamadas realizadas.
 - N°. de usuarios distintos.
 - Información del día 26 de Noviembre de 2009.

Estadísticas de Uso

➤ Ejemplo de estadísticas de uso:

- Picos de hasta 19 llamadas en una hora (entre las 3h y las 4h).
- Diferentes usuarios: 8 (entre las 3h y las 4h).
- Total de llamadas en el día: 134.
- Máxima duración media de llamada: 203 seg. (entre las 17h y las 18h).



■ Inicio de llamada ■ Llamantes

■ Duración media (seg.)

- ➔ **Objetivos**
- ➔ Sistemas de Diálogo
- ➔ Sistemas de Diálogo Orales
- ➔ El estándar VoiceXML
- ➔ **Aplicación Realizada**
- ➔ Demostración Práctica y Estadísticas de Uso
- ➔ **Conclusiones**
- ➔ Ruegos y preguntas

Conclusiones

- Se ha mostrado la utilidad de los sistemas de diálogo orales.
- Se ha realizado una aplicación útil para un determinado grupo de usuarios.
- Aplicación flexible y potente (amplias posibilidades de cara al usuario usando un código sencillo).
- Se ha mostrado la utilidad del estándar VoiceXML así como otros sistemas y herramientas en el desarrollo de la aplicación (web servers, Java, BD, etc.)

Conclusiones

- Mejoras y trabajo futuro:
 - Nuevos idiomas.
 - Actualización automática de la información en la base de Datos.
 - Ficheros de audio.
 - Recuperación y análisis automático de ficheros de log.

- ➔ Objetivos
- ➔ Sistemas de Diálogo
- ➔ Sistemas de Diálogo Orales
- ➔ El estándar VoiceXML
- ➔ Aplicación Realizada
- ➔ Demostración Práctica
- ➔ Conclusiones
- ➔ Ruegos y preguntas