

Modeling Voice Conversation Assistant

Maribella Rosas Perez

Juan Manuel Torres Castillo

IIT/UACJ

Ave. Del Charro 450 Norte

Cd. Juarez, Chihuahua

A1106928@alumnos.uacj.mx

Abstract. The present thesis aims at modeling of natural language within the framework of virtual assistants on mobile devices. Current mobile devices have a voice assistant, which is generated by a computer program that is able to recognize basic natural language that simulates a conversation. They can give information and offer a service through voice to text, text to voice.

Voice assistants that currently exist handle the processing of natural language. However, they do not recognize voice in some cases. You can find this kind of technology on mobile devices, computers, cars, etc., and anyone with the help of a smart mobile device can access it.

It should be mentioned that there are applications such as google Search, GPS navigation systems and more on mobile devices that work through a voice assistant and serve as support for people with disabilities. They have been created in different parts of the world, allowing them to be coupled with a simple simpler in daily tasks as well as in the workplace. However, they are little practical due to their general broad approach. They can confuse contexts.

In this work, we propose to develop a prototype of a voice assistant modeling, by processing the natural language, so that the user can have a smooth conversation with the assistant.

Sinopsis. La presente tesis tiene como objetivo el modelado del lenguaje natural dentro del marco de los asistentes virtuales en dispositivos móviles, ya que se tiene que un asistente de voz, el cual es generado mediante un programa informático que es capaz de reconocer de forma básica un lenguaje natural que simula una conversación para dar información y ofrecer un servicio mediante la voz a texto, texto a voz.

Los asistentes de voz que existen en la actualidad manejan el procesamiento del lenguaje natural, sin embargo, no lo reconocen favorablemente. Se pueden encontrar este tipo de tecnologías en dispositivos móviles, computadoras, autos,

etc., y cualquier persona con ayuda de un dispositivo móvil inteligente puede tener acceso.

Cabe mencionar que existen aplicaciones como google Search, sistemas de navegación GPS y otras más en dispositivos móviles que funcionan mediante un asistente de voz y sirven de apoyo para las personas que cuentan con alguna discapacidad, las cuales se han venido creando en diferentes partes del mundo, permitiendo que puedan acoplarse de una manera más sencilla tanto en las labores cotidianas como también en el ámbito laboral. Sin embargo, son poco prácticos debido a su amplio y general enfoque, que muchas veces lo confunde y hace que mezcle contextos o que se pierda en la conversación.

En este trabajo proponemos desarrollar un prototipo de un modelado de asistente de voz, mediante el procesamiento del lenguaje natural, esto con el fin que el usuario pueda mantener una conversación fluida con el asistente.

Planteamiento del problema

Los asistentes móviles se encontraron dificultades como calidad de síntesis, rapidez o facilidad de operación al momento de buscar algo, ya que estos no entienden de una manera fluida el lenguaje formal que las personas usan, las distintas formas de expresar o pedir algo.

Debido a las dificultades encontradas en los diferentes asistentes comerciales, se propone un estudio que mejore el uso del procesamiento del lenguaje natural para un asistente de voz, con el fin de tener una conversación fluida entre el asistente y el usuario.

Antecedentes

Siri

Fue creada en diciembre de 2007 por Dag Kittlaus (CEO), Adam Cheyer (VP Engineering) y Tom Gruber (CTO/VP Design) junto a Norman Winarsky del grupo SRI venture group y adquirida por Apple Inc. El 28 de abril del 2010, fue el primer asistente virtual disponible para los usuarios de teléfonos inteligentes que utiliza el procesamiento del lenguaje natural para responder preguntas.

Sherpa

Creado en español y su base de datos, una característica que le permite distinguir las palabras a la perfección y brinda respuesta correcta al instante. La plataforma está basada en la tecnología de reconocimiento de voz, lenguaje natural y la semántica, permitiendo diálogos naturales entre humanos y máquinas.

Sus ventajas, es que permite consulta información, enviar mensaje SMS o hacer llamadas, escuchar canciones, escribir comentarios en las redes sociales como Facebook o twitter, utilizan un modelo de probabilidad.

Su desventaja, es que el usuario realizar una petición, se considera que el asistente confunde las ordenes o da una respuesta fallida para aquellas solicitudes que el sistema no reconoce por fallas de conexión o por no tener claras las peticiones.

Los asistentes virtuales mencionados anteriormente, están más enfocados a buscar la información en las redes sociales, ya que si por ejemplo, en una oración al intercalar el orden de la frase de una manera diferente, estos realizan la indagación en la web y no existe una interacción directa con el usuario, ya que como se mencionó anteriormente, es tanta la información que existe en la red que puede causar una confusión que busca canciones o datos que pudieran tener un a fin con alguna palabra que de la frase.

Cortana

Otro asistente que fue lanzado en Estados Unidos y posteriormente en el 2014 estaba disponible es el resto de los países. Este asistente es específico de cada país, está adaptado a los patrones cotidianos de cada región del usuario, puede ser usado en diversos dispositivos, se activa con el comando “Hola Cortana”, para poder usar Cortana en español, se tiene que cambiar la configuración del sistema.

Cortana requiere que el usuario ingrese los comandos de lo que requiera de una manera clara, este asistente busca la información en la web o en su caso la información que busca será errónea y fuera de contexto, tendrá la repetir la orden incluso varias veces. Si existe un problema de conexión no realizará la orden hasta después de repetir incluso 10 veces lo orden.

Sistemas de ayuda al conductor (SAC).

La industria automotriz en su afán de ofrecerle al cliente lo mejor y esforzándose en la competitividad en el mercado de los sistemas de ayuda al conductor conocidos como SAC. La tarea principal de los sistemas es ayudar al conductor inválido a resolver cualquier dificultad que tenga mientras este conduciendo, como son, reconocimiento de voz y sistema de ayuda de navegación GPS. Uno de los fabricantes de automóviles de lujo anuncio Android a partir del 2015.

En el ejemplo del asistente de automóviles, estos manejan un lenguaje más técnico, las órdenes son estrictamente en un orden, sino reciben, el comando correcto, esto no se realiza.

Cuentan con una pantalla táctil a color en tablero de mandos situados delante del asiento del conductor muy parecido a una tableta, con funciones limitadas, dependiendo del modelo del auto este implica la cantidad de funciones y tareas.

Las funciones que se pueden realizar son: buscar una canción, responder el teléfono, introducir una ruta alternativa en el navegador GPS, realizar una búsqueda en internet, recibir un mensaje de texto, un correo electrónico o un mensaje de twitter.

Como se presentó, estos sistemas fueron creados para brindar seguridad a los conductores, como también, una competencia abierta entre las distintas empresas que se dedican a la fabricación de estos asistentes y en el avance de las diferentes tecnologías.

Criticando estos asistentes, se encuentra que estos brindan apoyo, pero no solucionan todos los problemas que pueden presentarse como por ejemplo, una dirección mal escrita.

Cabe mencionar que estos asistentes funcionan en automóviles electrónicos, previamente el usuario conectara el bluetooth y registrar el teléfono (marca y modelo), en los autos austeros, estos no traen asistente, pero si, se puede conectar el bluetooth y funciona por medio de un botón colocado en el volante, esto con el fin de poder contestar el teléfono.

Objetivos

Implementar un prototipo de un asistente de voz que realice procesamiento del lenguaje natural al escuchar vía voz.

Localizar objetos utilizando un conjunto de eventos realizados que se van a ir registrando en una base de datos piloto, de tal manera que el asistente pueda deducir el posible lugar donde se encuentre.

Diseñar una base de datos piloto que guardará los últimos registros.

Comprobación de resultados del prototipo con otros asistentes comerciales, para poder así, medir los resultados esperados y su procesamiento del lenguaje natural.

Diseñar una gramática piloto adecuada a un contexto definido.

Justificación

Con la creación de esta aplicación móvil desarrollada en un sistema operativo para un dispositivo inteligente, se espera que con el procesamiento del lenguaje natural el asistente pueda dialogar de una manera más fluida con el usuario sin salirse del contexto.

El proceso de preguntarle al dispositivo sobre la localización de un artículo dentro de un ambiente conocido como el hogar, por eventos almacenados, contribuirá que este asistente de voz pueda facilitar la búsqueda de objetos, haciéndola de una manera eficaz y personalizada.

Esta búsqueda ayudará a las personas que tienen problemas sobre recordar objetos dentro de un ambiente conocido como el hogar, facilitando la ubicación de objetos, como por ejemplo las llaves, medicinas, etc.

Marco teórico

Uno de los métodos en los asistentes de voz, es la inteligencia artificial, una de las definiciones de la inteligencia la especifica como <<desarrollo y utilización de ordenadores con los que se intenta reproducir los procesos de la inteligencia humana>> [1]

Gramática

Una gramática (G) es definida como un conjunto finito de reglas que describen toda la secuencia de símbolos pertenecientes a un lenguaje específico (L) y se considera una estructura algebraica, que está formada por cuatro elementos fundamentales:

$$G \{NT, T, S, P\}$$

NT es el conjunto de elementos No Terminales

T es el conjunto de elementos Terminales

S es el Símbolo inicial de la gramática

P es el conjunto de Reglas de Producción

Tipos de gramáticas según Chomsky

Gramáticas Tipo 0 (sin restricciones, recursivas). Incluyen todas las gramáticas formales. Generan todos los lenguajes que pueden ser reconocidos por una máquina de Turing (útiles para entender los procesos lógicos de las maquinas)

Gramáticas Tipo 1 (dependientes de contexto). Generan los lenguajes dependientes de contexto. Contienen reglas de producción de la forma.

$$\alpha A \beta \rightarrow \alpha \gamma \beta$$

A es un no terminal

α , β y γ son cadenas de terminales y no terminales.

α y β pueden ser vacíos, pero γ ha de ser distinto del vacío.

Se denominan gramáticas dependientes del contexto, porque, como se observa, A puede ser sustituido por γ si está acompañada de α por la izquierda y de β por la derecha.

Gramáticas Tipo 2 (independientes de contexto, libre de contexto). Generan los lenguajes libres de contexto. Están definidas por reglas de la forma.

$$A \rightarrow \gamma$$

A es un no terminal

γ Es una cadena de terminales y no terminales.

Se denominan independientes de contexto porque A puede sustituirse por γ independientemente de las cadenas por las que este acompañada. Los lenguajes independientes de contexto constituyen la base teórica para la sintaxis de la mayoría de los lenguajes de programación. Definen la sintaxis de las declaraciones, las proposiciones, las expresiones.

Gramáticas Tipo 3 (gramáticas regulares). Generan los lenguajes regulares. Las reglas se restringen a un único no terminal en la parte izquierda y una parte derecha compuesta por un único terminal que puede estar seguido o no de un único no terminal.



Fig. 1. Jerarquía de Chomsky

Según Chomsky, planteaba que los seres humanos nacen con la capacidad comunicarse con los demás, aun incluso, cuando nadie les hable, esto como parte de una formación biológica genéticamente de cada individuo y que al momento de nacer este se desarrolla a partir del nacimiento, donde cada uno desarrolla una gramática particular con la ayuda de sus padres.

Marco Tecnológico

En esta parte presentará los diferentes medios tecnológicos por los cuales se conseguiría llevar a cabo El Modelado de Conversación para Asistente de Voz. En primer lugar el sistema operativo “Android” que utilizaremos para llevar a cabo la aplicación móvil en dispositivos inteligentes.

Gramáticas de libre contexto.

Denominadas también gramáticas independientes del contexto, son el conjunto de reglas capaces de generar todas las posibilidades combinatorias de un lenguaje. Estos lenguajes libres del contexto son aquellos que pueden ser reconocidos por un autómeta de pila determinístico o no determinístico. La sintaxis de la mayoría de lenguajes de programación está definida mediante gramáticas libres de contexto.

Arboles de derivación

El árbol de derivación agrupa los símbolos de una cadena terminal en subcadenas y cuando se emplea en un compilador, es la estructura de datos que representan el programa fuente, que facilita la traducción del programa fuente a código ejecutable permitiendo que el proceso de traducción sea realizado por funciones naturales recursivas. Las derivaciones han demostrado ser extremadamente útil.

Materiales

Para diseñar el prototipo de Modeling Voice Conversation Assistant, se utilizará un asistente operativo de código abierto, una base de datos piloto, el Api para convertir la voz a texto.

Android Studio, por ser un software gratuito, ofrece herramientas para el diseño de aplicaciones, el cual puede ser utilizado en cualquier equipo con sistema operativo android, como puede ser una tablet, un dispositivo móvil e incluso para computadoras personales con ambiente Windows y Mac.

APIs para convertir la voz a texto

API Google, esta aplicación facilitará la tarea de convertir la voz a texto, que se necesitará para poder llevar a cabo la tesis, que ayudará en el procesamiento del lenguaje natural.

W3C-SRGS (Speech Recognition Grammar Specification)

En la actualidad W3C trabaja con estándares como VoiceXML, SSML (Speech Synthesis Markup Language), SRGS (Speech Grammar Recognition Specification), SISR (Semantic Interpretation for Speech Recognition), In IML (Ink Markup Language) y herramientas de integración de datos de entrada como EMMA (Extensible MultiModal Annotation Markup).

SSML (Lenguaje de Marcado de Síntesis del Habla), forma parte del proceso de salida de información y funciona como componente de conversión ayudando a gene-

rar habla sintética. Este lenguaje tiene como principal objetivo ofrecer a los autores de contenido un camino estándar para controlar aspectos del habla como puede ser la pronunciación, volumen, tono y velocidad.

Lenguaje de programación Android Studio

Este es un software oficial para desarrollar aplicaciones para Android, el cual utiliza una licencia libre, programado en java y es multiplataforma, presentado por Google el 16 de Mayo del 2013 en el congreso de desarrolladores Google I/O, con el único objetivo de crear un entorno dedicado a la programación de aplicaciones para dispositivos android. Requisitos del sistema.

- Mínimo 2 GB de RAM, recomendado 4 GB de RAM
- 400 MB de espacio en disco
- Resolución de pantalla 1280 x 800 mínimo

Este sistema fue escogido ya que ayuda a mejorar el asistente de voz por ser adaptable a diferentes pantallas y resoluciones, interactúa con librerías mediante JNI (Java Native Interface).

Alcances y Limitaciones

En esta sección se enlistan las delimitaciones en tiempo y espacio, así como también las limitaciones que afectan directamente en el plan y al mismo tiempo el desarrollo del proyecto del prototipo Modelado de Asistente de Voz.

Alcances

- El asistente contará con una base de datos que registrará los diferentes eventos que el usuario realice.
- Podrá llevar a cabo diálogos pequeños con el usuario, vía voz.
- Se utilizará un sistema operativo de código abierto
- Un subconjunto del lenguaje español, adecuado a un ámbito definido.

Limitaciones

- Ésta funcionará dentro de un contexto conocido por ejemplo “el hogar”.
- El tiempo para el desarrollo de este proyecto es de 8 meses
- Se utilizarán herramientas de código abierto.
- Se usará una base de datos piloto, que no necesariamente corresponda con la realidad.

- Se diseñará una gramática piloto inicial, no sofisticada para comenzar a resolver algunos de los problemas de contexto

Diagrama

El siguiente diagrama es un modelo de la forma en la cual el prototipo funcionará, se observa como el usuario vía voz y mediante el api de google, emite una orden al asistente de voz, entonces mediante una descomposición de la gramática, el texto es descompuesto para pasar a una semántica para definir una acción y poderla ejecutar. Si no ejecuta, entonces el asistente mediante un dialogo corto, le dice al usuario que no entendió la orden y necesita de nuevo la orden para poderla ejecutar.

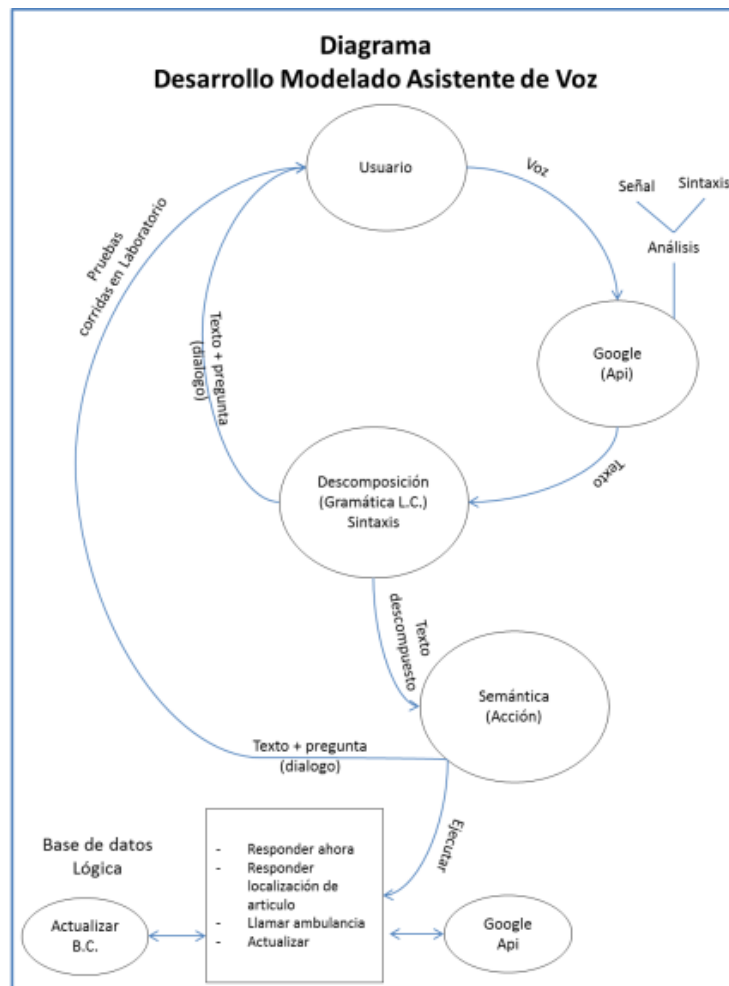


Fig. 2. Diagrama modelado

Ejemplos de re-estructuración de frases

En estos dos ejemplos que se presentan de la re-estructuración de frases, se observa como está formada una oración, la estructura que existe en cada una de ellas, en el lenguaje español las oraciones no están simplemente formadas por palabras, sino, que estas tienen la estructura sintáctica de las oraciones.

Me duele la cabeza	Pronombre	Me
	Artículo	La
	Verbo	Duele
	Nombre	Cabeza

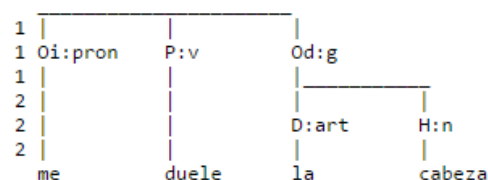


Table 1. Primer ejemplo de la descomposicion de frases

Donde están mis medicinas	Adverbio	Donde
	Verbo	están
	Pronombre	Mis
	Nombre	medicinas

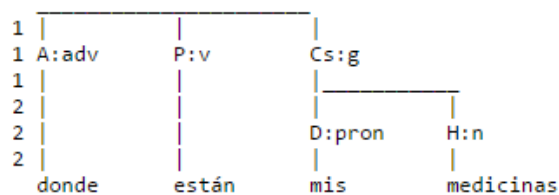


Table 2. Segundo ejemplo de la descomposicion de frases



Fig. 3. Ejemplo árbol de la frase “donde están mis medicinas”

Escenario de frases propuesta para las pruebas de la aplicación

Oraciones

Preguntas	¿Qué hora es?	
	¿Dónde están mis medicinas?	El asistente le contestara al usuario la última ubicación de las medicinas, previamente registradas por el usuario
	¿Cuál es el horario de mis medicamentos?	El asistente responderá el horario registrado de los medicamentos, previamente guardados en la aplicación
Frases	Me duele la cabeza	El asistente posiblemente pueda sugerir algún medicamento para el dolor

Órdenes	Me siento mal	El asistente pueda interactuar con el usuario, preguntándole, ¿deseas que llame al doctor?
	Llama a mi mama	El asistente realizara la llamada
	Envía un mensaje al usuario.	El asistente enviara el mensaje
	Llama a la ambulancia	El asistente realizara la llamada al número de emergencia registrado previamente por el usuario

Conclusiones

Este prototipo con la ayuda de Android Studio pretende mediante el procesamiento del lenguaje natural crear un prototipo que interactúe con la ayuda de una aplicación a que un usuario pueda localizar objetos sobre eventos previamente registrados.

Cabe señalar que se usó un vocabulario en un ambiente conocido “una casa”, esta aplicación trabajara en conjunto con una base de datos lógica, en la cual, se registrarán eventos que deberán de ser registrados por el usuario.

Por la complejidad del lenguaje natural se considera continuar modelando el prototipo para automatizar la base de conocimientos.

También, realizar pruebas futuras en diferentes escenarios, como por ejemplo el trabajo, para que este puede adaptarse y ser utilizado.

Agradecimientos

A Dios, por habernos permitido vivir y afrontar las dificultades, por darnos salud y el conocimiento para lograr y poder llevar a cabo nuestros objetivos, por su bondad y amor.

Por el apoyo y motivación que recibimos de nuestros familiares y amigos en el caminar de la carrera profesional y terminación de la tesis, por impulsarnos a seguir adelante.

Y a aquellos amigos que ya no están y que fueron ejemplos de constancia, fortaleza de vida y de superación en el transcurso de nuestra vida, que han dejado una huella muy importante.

Especialmente al maestro Dr. Juan Carlos Acosta Guadarrama, por sus consejos, apoyo y conocimientos transmitidos para la realización de la tesis.

Referencias

- [1] (2016) SpeechRecognizer. [Online].
<http://developer.android.com/intl/es/reference/android/speech/SpeechRecognizer.html>
- [2] (2016, Julio) La importancia de los Sistemas de Ayuda al Conductor (SAC). [Online]. <http://coches1a.es/la-importancia-de-los-sistemas-de-ayuda-al-conductor-sac/>
- [3] George Coulouris, Jean Dollimore, Tim Kindberg,.: Pearson Education Limited, 2005.
- [4] Roman Martinez/Elda Quiroga, *Estructuras de datos, Referencia practica con orientacion a objetos*. Monterrey, N.L. Mexico: Thomson Learning, 2001.
- [5] Luis Joyanes Aguilar, *Fundamentos de programacion, algoritmos, Estructuras de datos y Objetos*. Aravaca, Madrid: Mc Graw Hill, 2003.
- [6] Julio Carmona Alonso, *Base de Datos*. Sevilla, 2000.
- [7] Alejandro P. Ayala, *Lenguaje Natural*. Mexico: Instituto Nacional Politecnico, 2006.
- [8] Jr., David E Penney C.H. Edwards, *Ecuaciones Direccionales Elementales*.: Pearson Educacion, 2000.
- [9] Fernando Arango Isaza Carlos Mario Zapata Jaramillo, "Los modelos verbales en Lenguaje Natural y su utilización en la elaboración de esquemas conceptuales para el desarrollo de software: una revisión

crítica," *REVISTA Universidad EAFIT*, 2005.

- [10] Arboleda Carolina, Garcia Eliana, Posada Alejandro, and Torres Robinson, *Diseño y construcción de un Prototipo de Interfaz Cerebro-Computador para facilitar la Comunicación de Personas con Discapacidad Motora*. Medellin, Colombia: 2009, 2009.
- [11] Jose Guillermo Carrillo. (Mayo, 2004) Fundacion Internacional. [Online].
http://www.fundacionjoseguillermocarrillo.com/sitio/testimonio_noam_chomsky.php
- [12] Lucho Castillo. (2012, Noviembre) Android. [Online].
<https://androidos.readthedocs.org/en/latest/data/glosario/#sqlite>
- [13] Javier Lopez Garcia, *Herramienta movil para el guiado de personas con discapacidad cognitiva en entornos Interiores*. Universidad Autonoma de Madrid: 2015, 2015.
- [14] Arturo Garcia Gonzalez, *Ecuaciones Diferenciales*. Cd. Juarez, Chih.: UACJ, 2000.
- [15] Pablo Ibáñez. (2014, Febrero) Los sistemas de control por voz en los coches para mejorar la seguridad. [Online].
<http://www.circulaseguro.com/los-sistemas-de-control-por-voz-en-los-coches-para-mejorar-la-seguridad/>
- [16] Lee Hetheringto, Ezra Story Johan Schalkwyk, "Speech Recognition with Dynamic Grammars Using Finite-State Transducers," GENEVA, 2003.
- [17] Bernal Zamora Leonardo and Salamanca Oscar Miguel, "*Manos Que Hablan*". *Prototipo de Aplicación en Android para el Aprendizaje del Alfabeto Dactilológico Para Colombia*.
- [18] Samaniego Pilar, Laitamo Sanna-Mari, Valerio Estela, and Francisco Cristina, *Informe sobre el uso de las Tecnologías de Información y Comunicación (TIC) en la Educación para Personas con Discapacidad*.

Quito, Ecuador: 2012, 2012.

- [19] Nacho Saenz. (2012, Octubre) Revista Byte. [Online].
<http://www.revistabyte.es/movilidad/10nace-sherpa-primer-asistente-de-voz-en-espanol-que-entiende-el-lenguaje-humano-y-maneja-el-telefono-sin-necesidad-de-teclear012/>

- [20] Moviles Tecnologia. (2014, Diciembre) Blog de tecnología, con todas las noticias de móviles, celulares y smartphones especialmente, tratando Android, Apple y Windows Phone. [Online]. http://smartphones-celulares-moviles.blogspot.mx/2014_06_01_archive.html