Procesamiento del Lenguaje Natural, Revista nº 59, septiembre de 2017, pp. 149-152 recibido 03-04-2017 revisado 12-05-2017 aceptado 18-05-2017

IXHEALTH: Un sistema avanzado de reconocimiento del habla para la interacción con sistemas de información de sanidad

IXHEALTH: An advanced speech recognition system to interact with healthcare information systems

Pedro José Vivancos-Vicente[1], Juan Salvador Castejón-Garrido[1], Mario Andrés Paredes-Valverde[2], María del Pilar Salas-Zárate[2], Rafael Valencia-García[2]
[1] VOCALI SISTEMAS INTELIGENTES S.L.
Parque Científico de Murcia. Ctra. de Madrid km. 388. Complejo de Espinardo. 30100 Murcia. Spain
pedro.vivancos@vocali.net, juans.castejon@vocali.net
[2] Universidad de Murcia
Facultad de Informática Campus de Espinardo, 30100, Murcia, España
marioandres.paredes@um.es, mariapilar.salas@um.es, valencia@um.es

Resumen: El objetivo del proyecto IXHEALTH es desarrollar una plataforma multilingüe basada en reconocimiento del habla que permita a profesionales de la salud llevar a cabo tareas tales como la redacción de informes médicos, así como interactuar con sistemas de información sanitarios mediante comandos de voz. Todo ello, bajo un mecanismo de seguridad basado en biometría de voz que evite que personas no autorizadas editen información sensible gestionada por este tipo de sistemas. Este proyecto ha sido desarrollado por la empresa VOCALI en conjunto con el grupo de investigación TECNOMOD de la Universidad de Murcia, y financiado por el Instituto de Fomento de la Región de Murcia.
Palabras clave: Reconocimiento del habla, biometría de voz, sistemas de información sanitarios

Abstract: The IXHEALTH project aims to develop a multilingual platform based on speech recognition that allows healthcare professionals to perform transcription and dictation activities for the generation of medical reports, as well as to interact with healthcare information systems by means of voice commands. These tasks are performed through a biometric voice-based security mechanism that avoids non-allowed users to edit sensitive data managed by this kind of systems. This project has been developed by the VOCALI enterprise in conjunction with the TECNOMOD research group from the University of Murcia, and it has been founded by the Institute of Promotion from the Region of Murcia.
Keywords: Speech recognition, biometric voice, healthcare information systems

1 Introducción y objetivos del proyecto

Los sistemas de reconocimiento del habla están cada vez más presentes en la sociedad. Concretamente, la sanidad es uno de los dominios donde estos sistemas son imprescindibles para la redacción de informes y transcripción en diversas especialidades tales como radiología (Akhtar, Ali, y Mirza 2011) y patología (Al-Aynati y Chorneyko 2003). Sin embargo, el personal sanitario demanda mejores funcionalidades que les permitan manejar los sistemas de información a través del habla y no mediante interfaces tradicionales (teclado, mouse, pantalla táctil) ya que algunos médicos por su labor no pueden manejar un ordenador o tener la vista en la pantalla mientras trabajan.

Por otro lado, en sistemas de información de sanidad es crucial la incorporación de mecanismos de seguridad tal como la biometría

de voz ya que los informes médicos tienen responsabilidad jurídica. De esta manera, si algún profesional sanitario efectúa un mal diagnóstico puede tener responsabilidades y por eso es muy importante que quede registrado quién realizó los informes. Además, con la biometría de voz se puede comprobar que el usuario que hace el informe es quien dice ser, ya que es una práctica habitual que un médico no cierre la sesión del ordenador y otra persona pueda utilizar el sistema pudiendo alterar información sensible.

El reto principal de este proyecto es desarrollar un sistema avanzado de reconocimiento del habla en lenguaje natural que permita la definición y gestión de comandos de voz para interactuar con sistemas de información de sanidad. Todo esto complementado con biometría de voz (validación del usuario en tiempo real) y soporte para múltiples idiomas, concretamente español y portugués. De esta manera, el sistema permitirá a profesionales de la salud agilizar sus tareas y con ello incrementar su productividad.

2 Estado actual del proyecto

Actualmente, el sistema se encuentra en fase de validación en centros sanitarios. De manera general, IXHEALTH es una plataforma multilingüe para el reconocimiento avanzado del habla que permite a profesionales de la salud realizar actividades de transcripción y dictado para la edición de documentos clínicos y el llenado de formularios electrónicos, así como definir y gestionar comandos de voz para interactuar con sistemas de información de salud, incluyendo el sistema operativo en el que se ejecutan. En la siguiente sección se describe la arquitectura de la plataforma.

2.1 Arquitectura de la plataforma IXHEALTH

Como se aprecia en la Figura 1, el sistema IXHEALTH se compone de cinco módulos: (1) módulo de reconocimiento del habla, el cual permite a los usuarios interactuar con sistemas de información a través de comandos de voz en lenguaje natural, así como realizar actividades de transcripción y dictado tales como la edición de documentos clínicos; (2) síntesis de voz, este permite a la plataforma IXHEALTH leer texto contenido en los sistemas de información y convertirlo en voz, permitiendo así a los profesionales de la salud realizar otras actividades sin prestar atención a la interfaz principal; (3) anotación semántica, este obtiene una interpretación semántica de la información involucrada en el proceso de reconocimiento de voz, como registros médicos, informes de pruebas médicas y ensayos clínicos, entre otros (4) biometría de voz, este realiza una verificación del usuario en tiempo real para evitar el uso de sistemas de información de sanidad por usuarios no autorizados. y (5) gestión de recursos lingüísticos multilingües, el cual permite la gestión de comandos y recursos lingüísticos, utilizados por otros módulos, para idiomas tales como el portugués y el español.

A continuación, se describen los módulos mencionados anteriormente.

Figura 1: Arquitectura de IXHEALTH

2.2 Administración de recursos lingüísticos multilingüe

Este módulo permite gestionar los recursos lingüísticos utilizados por el módulo de reconocimiento del habla de tal manera que la edición de estos no afecte el desempeño global del sistema. Los recursos administrados se describen a continuación.

Modelo acústico. Este provee una representación estadística de la relación entre una señal de audio y los fonemas y otras unidades lingüísticas que componen el habla. Este modelo está basado en el Modelo Oculto de Markov (Juang y Rabiner 1991).

Modelo del lenguaje. Este determina la función de probabilidad conjunta de secuencias de palabras en un lenguaje. Este modelo se basa en un corpus de textos que se usa para calcular la probabilidad de que una determinada palabra aparezca antes o después de otra.

Diccionarios. Estos contienen términos específicos del dominio incluyendo su

respectiva pronunciación la cual se basa en fonemas. Estos recursos representan un componente importante del sistema ya que, en el ámbito sanitario, cada especialidad médica tiene un vocabulario específico cuya detección mejora el desempeño general del sistema.

Gramática de comandos. Este componente representa todas las formas posibles de hacer referencia a un comando específico definido por el usuario, incluyendo sus sinónimos. La definición de estas gramáticas se basa en SRGS (Speech Recognition Grammar Specification) (Hunt y McGlashan 2004) un estándar que proporciona un alto nivel de expresividad de una gramática libre de contexto.

La plataforma IXHEALTH provee soporte para el español y portugués por lo que existen modelos acústicos, modelos de lenguaje, diccionarios y gramática de comandos para cada lenguaje.

2.3 Reconocimiento del habla

Este módulo integra un motor de dictado y un motor de comandos de voz. El primero de ellos permite al usuario llevar a cabo tareas de dictado y transcripción tal como la edición de documentos clínicos. El segundo detecta comandos específicos para ser ejecutados por el sistema de información o el sistema operativo sobre el que se ejecuta. Dichos motores comparten el mismo reconocedor del habla, por lo que ambos funcionan en paralelo. De esta manera, cuando el reconocedor del habla recibe la señal de voz, ambos motores analizan la señal para determinar si el usuario ha provisto un comando predefinido o si desea llevar a cabo tareas de dictado. Cabe mencionar que el sistema prioriza a los comandos de voz.

El motor de comandos de voz reconoce dos tipos de comandos: (1) comandos simples, el cual consiste en una secuencia de invocaciones fijas. Un ejemplo de este tipo de comandos es "iniciar dictado"; y (2) comando de dos partes, el cual contiene una secuencia de invocaciones fijas y un parámetro que consta de una o más palabras. Un ejemplo de este tipo de comando es "selecciona introducción", donde "selecciona" representa el comando, e "introducción" representa el parámetro, en este caso, la sección a ser seleccionada.

2.4 Biometría de voz

Este módulo implementa un mecanismo de biometría de voz en tiempo real que permite autenticar al usuario, es decir, asegurar que un usuario es quien dice ser. Previo al proceso de autenticación, este módulo genera una huella de voz por cada usuario. Esta huella es comparada con la señal de voz recibida por el usuario en tiempo real. Los resultados de la comparación se cuantifican y se comparan con un umbral de aceptación/rechazo para determinar si las dos huellas son suficientemente similares para que el sistema acepte la identidad. Esta decisión se basa en una puntuación LLR (log-likelihood ratio).

2.5 Síntesis de voz

Este módulo permite a la plataforma leer texto contenido en los sistemas de información y convertirlo en voz. De esta manera, es posible que los profesionales de la salud realicen otras actividades sin prestar atención a la interfaz gráfica, ahorrando así tiempo y esfuerzo. Este proceso se basa en una transcripción de grafema a fonema de las oraciones a pronunciar para lo cual lleva a cabo cinco pasos principales: (1) organiza las frases de entrada en una lista manejable de palabras, (2) realiza un proceso LTS (Letter-To-Sound) con el fin de determinar la transcripción fonética del texto entrante; (3) identifica las propiedades de la señal de voz relacionadas con los cambios audibles en tono, volumen y longitud de la sílaba con el fin de generar una estructura sintáctica-prosódica, (4), produce un bloque de concatenación de segmentos de voz, es decir, transiciones fonéticas y coarticulaciones, utilizadas como unidades acústicas finales, y (5) genera una única señal compacta que contiene todos los segmentos de voz de forma coherente. Esta señal se almacena como un archivo mp3 para que cualquier dispositivo pueda reproducirlo.

2.6 Anotación semántica

Las tecnologías semánticas proporcionan una base consistente y confiable que puede ser utilizada para enfrentar los desafíos relacionados con la organización, manipulación y visualización de datos y conocimientos. Por lo tanto, este módulo realiza la anotación semántica de los recursos involucrados en los sistemas de información sanitaria tales como registros médicos, informes de pruebas médicas y ensayos clínicos, entre otros, con el fin de obtener una interpretación semántica de los mismos. Este módulo se basa en trabajos previos del grupo de investigación

TECNOMOD (Paredes-Valverde et al. 2015) y consta de dos fases principales.

En primer lugar, se encuentra la fase de pre-procesamiento de texto, la cual realiza el proceso de tokenización, división de oraciones y stemming. Este último se refiere a reducir las palabras a su raíz. En segundo lugar, se lleva a cabo la fase de detección de conceptos médicos, la cual detecta y anota los conceptos médicos contenidos en el texto de entrada. Estos conceptos se identifican mediante reglas JAPE y gazetteers. Por un lado, JAPE es un mecanismo de reglas basado en expresiones regulares que permite el reconocimiento de expresiones sobre anotaciones realizadas en documentos. Por otra parte, un gazetteer consiste en un conjunto de listas que contienen nombres de entidades tales como diagnósticos, procedimientos, alergias, alertas, entre otros. Con el objetivo de proporcionar interoperabilidad semántica, estos gazetteers se basan en las siguientes terminologías estándar.

SNOMED-CT. Es la terminología clínica multilingüe más completa del mundo. Contiene contenido clínico completo y científicamente validado que permite una representación consistente del contenido clínico en los registros de salud electrónicos.

CIE-9. La Clasificación Internacional de Enfermedades, novena edición, clasifica las enfermedades, las condiciones y las causas externas de las enfermedades y lesiones (mortalidad y morbilidad).

CIE-10. Representa la décima revisión de la Clasificación Internacional de Enfermedades presentada anteriormente.

CIAP-2. La Clasificación de Atención Primaria es una taxonomía de términos y expresiones comúnmente utilizados en medicina general. Recopila las razones de la consulta, los problemas de salud y los procesos de atención.

3 Trabajo a futuro

Independientemente de los resultados obtenidos de la fase de evaluación en la que se encuentra actualmente el sistema será necesario mejorar la precisión en el reconocimiento de palabras ya que, en el dominio de la salud, un error de reconocimiento de palabras puede cambiar el significado completo de un informe, creando problemas de salud de los pacientes, lo que incrementaría los costos de sanidad.

Con el fin de aumentar la precisión del reconocimiento del habla, se pretende prestar especial atención a la mejora del modelo de lenguaje (español y portugués). Además, planeamos realizar pruebas continuas de la plataforma a lo largo de fases incrementales. Cada fase contará con la participación de profesionales sanitarios de diferentes especialidades. Al final de cada fase se obtendrá la tasa de reconocimiento de palabras y se analizarán los resultados con el objetivo de detectar las principales causas de errores de reconocimiento de palabras, así como para medir el desempeño de nuestro sistema en diferentes especialidades. Finalmente, planeamos implementar la integración semántica de datos de ensayos clínicos, y proveer servicios como búsquedas semánticas en ensayos clínicos y datos de pacientes.

Agradecimientos

Este trabajo ha sido financiado por el Instituto de fomento de la Región de Murcia (Ref. 2015.08.ID+I.0011)

Bibliografía

Akhtar, W., A. Ali, y M. Kashif. 2011. Impact of a Voice Recognition System on Radiology Report Turnaround Time: Experience from a Non-English-Speaking South Asian Country, *AJR. American Journal of Roentgenology* 196 (4): W485; author reply 486. doi:10.2214/AJR.10.5426.

Al-Aynati, M., y K. Chorneyko. 2003. Comparison of Voice-Automated Transcription and Human Transcription in Generating Pathology Reports. *Archives of Pathology & Laboratory Medicine* 127 (6):721–25.

Hunt, A., y M. Scott. 2004. Speech Recognition Grammar Specification Version 1.0. W3C Recommendation, March.

Juang, B. H., and L. R. Rabiner. 1991. Hidden Markov Models for Speech Recognition. *Technometrics* 33 (3):251–72.

Paredes-Valverde, M., M. A. Rodríguez-García, A. Ruiz-Martínez, R. Valencia-García, and G. Alor-Hernández. 2015. ONLI: An Ontology-Based System for Querying DBpedia Using Natural Language Paradigm. *Expert Systems with Applications*. 42(12):5163–5176.