

Procesamiento del Habla – Trabajo Práctico 1

Departamento de Computación, FCEyN, UBA

Segundo cuatrimestre 2019

Implementación de un sistema de diálogo hablado

Este trabajo consiste en implementar un **sistema de diálogo hablado** (SDH) en un dominio a elección. Por SDH entendemos un sistema que interactúe con el usuario primariamente mediante el habla (tanto input como output).

Para las componentes de síntesis y reconocimiento, debe usarse software ya existente (ver abajo). El foco del TP debe estar en el funcionamiento global del sistema: diseño e implementación de las restantes componentes (administración del diálogo, y comprensión y generación del lenguaje), así como su articulación efectiva con las componentes de ASR y TTS.

Dominio

Para la elección del dominio del SDH sólo deben tenerse en cuenta dos requisitos: que la interacción involucre (i) reconocimiento del habla y (ii) síntesis del habla. Opcionalmente, también estaría bueno que incluya técnicas de procesamiento de señales para modificar los audios generados, o bien técnicas de aprendizaje automático para detectar características del usuario.

¡Usen su imaginación! Algunas ideas:

- Una variante de *Eliza*: <http://deixilabs.com/eliza.html>.
- Un sistema para proveer información de la web: resultados de deportes, estado del tiempo, horarios de medios transporte, artículos de diarios, posts de Twitter, etc.
- Un juego para transformar la voz del usuario: el sistema graba y reproduce lo que el usuario dice (al principio sin hacer nada). El usuario puede decir comandos para aplicar transformaciones al audio. Por ejemplo, al decir algo como “Comando: subir velocidad 10 %”, el sistema responde “Comando aceptado, subiendo velocidad 10 %” y a partir de entonces las cosas que el usuario diga (excepto comandos) se reproducen con mayor velocidad. Comandos posibles: subir/bajar tono/velocidad/volumen, invertir, aplicar filtros, etc.
- Un traductor automático, que pase a otro/s lenguaje/s lo que el usuario diga, usando alguna API gratuita de servicio de traducción en la nube.

Metáfora “presionar-para-hablar”

Por simplicidad, el sistema puede implementarse usando la metáfora “presionar-para-hablar” (*push-to-talk*), similar a la forma de mandar mensajes de audio en WhatsApp. Los scripts de ejemplo disponibles en el archivo `presionar-para-hablar.zip` tienen una implementación simple de esta metáfora en Python. Por supuesto, pueden implementar la interacción como prefieran. Otra opción interesante sería implementar un bot de Telegram, por ejemplo (<https://core.telegram.org/bots>).

Opciones gratuitas para las componentes de TTS y ASR:

- Pocketsphinx: <https://cmusphinx.github.io/wiki/tutorialpocketsphinx/>, con modelos acústicos en español: <https://sourceforge.net/projects/cmusphinx/files/AcousticandLanguageModels/Spanish/>
- Festival: <http://www.cstr.ed.ac.uk/projects/festival/>, con una voz en español: <https://sangonz.wordpress.com/2010/05/22/spanish-voices-for-festival/>
- API de IBM Watson, de uso gratuito para pruebas. El script `IBM_services.py` explica cómo usarlo. (Gracias a Ramiro Gálvez y Lara Gauder por la ayuda con esto.)
- Hay varias opciones más de servicios en la nube. Son libres de usar las que prefieran.

Modalidad de entrega

- El trabajo se debe realizar en grupos de **dos integrantes**.
- **Antes del 10 de noviembre** (y cuanto antes, mejor): Enviar una breve descripción del dominio elegido a `uba.habla@gmail.com`, para ver si está bien y poder continuar con el diseño y la implementación.
- El **miércoles 27 de noviembre** los autores deberán realizar una presentación oral y demostración del trabajo en clase, ya sea en una computadora personal o del laboratorio. Esta presentación durará 20 minutos, durante los cuales se deberá explicar el dominio, mostrar las funcionalidades del sistema y describir su implementación.
- Además, se deberá entregar un breve informe, en formato PDF y de cuatro (4) carillas como máximo, que describa los mismos aspectos de la presentación oral. Este informe (y solo el informe; no enviar código) debe enviarse a `uba.habla@gmail.com`. La fecha límite de entrega del informe es el **martes 26 de noviembre a las 23:55**. Los TPs entregados fuera de término serán aceptados, pero la demora incidirá negativamente en la nota.

Modo de evaluación

El trabajo se evaluará desde la óptica de la experiencia de usuario, valorando la robustez del sistema y la fluidez de la interacción. No se evaluarán aspectos estéticos de la interfaz, sino estrictamente los relativos al uso del habla para interactuar con el sistema.

El TP tiene una nota máxima de 100 puntos. Cumplir con los objetivos mínimos (es decir, un sistema que funcione y haga lo pedido) otorga 70 puntos. Los restantes 30 puntos corresponden a la *fluidez* y la *complejidad* de la interacción mediante el habla, según la opinión del profesor.