|  |  |  |
| --- | --- | --- |
| **Módulo 3 – Tema 5** | Jim Glass |  |

**Video 1:**

HCI:

* PC con Windows.
* Celulares con el touch.
* Smart watches.

Como nos vamos a comunicar con todos los dispositivos

* Probablemente no con el teclado y mouse.
* El habla va a ser una excelente modalidad.
* Siri, dispositivos Android, Cortana, Microsoft, Amazon echo.
* Más allá del habla, debe haber entendimiento.
* Diálogo.

Por qué?

* Natural. No necesitamos aprender.
* Flexible. Deja las manos, vista para hacer otras cosas.
* Eficiente para muchos escenarios. Si necesitamos escribir lo mismo!
* Económico.

Es ideal cuando:

* El espacio de información es amplio y complicado.
* Usuario inexperto.
* Los dispositivos son pequeños.

Funcionamiento **reconocimiento** de voz:

* Voz humana.
* Interpretación y traducción a texto en la CPU.

**Síntesis** de voz: viceversa también!

IMG en “**Reconocimiento-Entendimiento-Generación-Síntesis.png**”

Interfaces avanzadas:

* Modo de conversación.
* Entender el reconocimiento verbal.
* Verbalizar la respuesta.

**Video 2**

**Prototipos.**

Historia:

* 80s solo equipos locales.
* 90s: tecnología en la red. A fines se comenzó con sistemas gratuitos en el antiguo sistema de teléfonos. Nros. gratuitos para el clima, etc.
* 2000s: a gran escala.

Escenarios (comunes al as personas):

* Información de viaje: consultar, explorar desde app. en el teléfono. Basado en reconocimiento de voz-claves de texto-búsqueda.
* Nutrición: prototipo más recienta. El usuario pasa información de lo que come. Reconoce, interpreta y muestra la información nutricional de cada producto.

**Video 3**

**Tecnologías.**

Arquitectura del sistema de nutricional: “**Arquitectura sistema nutricional.png**”.

Tecnologías: **“Tecnologías.png”**.

Reconocimiento automático de voz: **“Reconocimiento de voz.png”** gran problema de búsqueda.

Semántica: identificación de conceptos, agrupamiento,

Contextualizar la interpretación o meaning: por ejemplo decir que tiene una copa de yogurt y luego del resultado decir yogurt griego => contexto más amplio.

Manejo del diálogo: gran reto. La mayoría de los procesadores de diálogo no han sido diseñados para público en general.

* Pre retrieval. Input ambiguo -> necesidad de clarificar para que sea una sola consulta.
* Post retrieval. Múltiple -> Respuesta única:

Generación de voz: si bien se puede dar resultados al usuario vía tabla, quizá se quiera informar por medio de voz.

**Video 4**

**Modalidades.** Por ejemplo para decir una contraseña no sería lo mejor. **“Modalidades.png”**.

Ejemplo de City Web desde tableta.

**Video 5**

**Modalidades distintas en paralelo SIMBIOSIS MULTIMODAL.**

Paralingüística: información no lingüística Contenido emocional.

Identificación multimodal de la personas utilizando voz y reconocimiento facial: cada uno tienen un 2% de error y se logró al combinarlos reducirlo a 0.2 %.

NO INVASIVOS.

**Video 6**

**Inconvenientes y oportunidades.**

Procesamiento baja potencia.

* Interacción con dispositivos de baja potencia.
* Algunas operaciones se pueden hacer localmente. Otra remotamente.
* Combinación VAD ASR local y remoto dependiendo del escenario **“Procesamiento baja potencia.png”**.
* En **“Procesamiento baja potencia2.png”** se puede observar que el aprendizaje se puede hacer primero y luego dicho componente (arriba) se puede comprimir listo para utilizarse.
* Se crearon circuitos de baja potencia (ultra baja potencia). Se podrían incorporar por ejemplo a fitbit y alimentar con temperatura del cuerpo, ni siquiera necesitarán batería.
* Se está trabajando con verificador de ID del hablante.

**Video 7**

**Retos.**

Iteración multimodal avanzada.

Prototipos:

* Ford Sync: las personas se están quejando. Necesita mejorar.
* Robots. Montacargas autónomo.
* No hablar a un dispositivo de voz para que este luego envíe instrucciones al dispositivo real, sino que hablar directamente, GRAN DESAFIO. **Amazon echo.**
* Otro reto es que las información no es estructurada y los sistemas convencionales si.

**Video 8**

Interpretación más amplia del contexto. Problemas de realizar las preguntas adecuadas.

Actualmente no pueden hacerlo los dispositivos.

**Interpretación contextual del propósito general. DEASFIO!!!!!**

Iteración en múltiples idiomas. **REPRESENTACION DEL SIGNIFICADO DEBE SER INDEPENDIENTE DEL IDIOMA. INTERLINGUA**

**Reflexiones:**

* Inevitable en IoT basarse en interfaces de voz.
* Deben ser robustas, diferentes ambientes, contextos, entender el contexto más amplio.
* Incorporar diferentes modalidades.
* Mili linguales.
* Actualmente en smatphones es la punta del iceberg, queda mucho por trabajar.