



UT.4 Introducción al procesamiento del lenguaje natural

Bloques de la UT4:



1. Procesamiento del lenguaje natural. (NLP)
2. Caso práctico: Asistente virtual en el hogar.
3. Desafíos.
4. Laboratorio: Ejemplos prácticos.

1. Procesamiento del lenguaje natural



El procesamiento del lenguaje natural (NLP por sus siglas en inglés) consiste en la habilidad de una máquina para procesar información comunicada mediante el uso del lenguaje natural.

Se crean modelos computacionales del lenguaje suficientemente detallados que permitan escribir programas informativos que realicen distintas órdenes o peticiones donde interviene el lenguaje natural.

1. Procesamiento del lenguaje natural



Se podría decir que el NLP consiste en usar una expresión natural que pueda tener comunicación con la computadora directamente, por medio escrito o comando de voz, facilitando las órdenes o peticiones con el lenguaje, o seguir desarrollando modelos que ayuden a la comprensión humana y sus mecanismos que se relacionan al lenguaje.

1. Procesamiento del lenguaje natural



Existen 2 áreas principales en el procesamiento del lenguaje natural

- TTS (Text To Speech) -> Generación de texto a voz.
- ASR (Automatic Speech Recognition)-> Reconocimiento de voz.

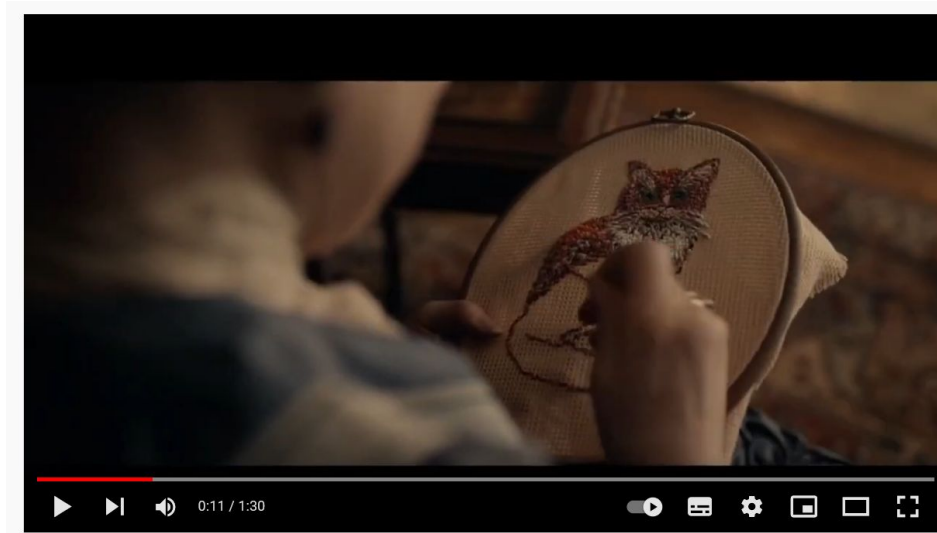
1. Procesamiento del lenguaje natural



Puntos clave en el procesamiento del lenguaje natural:

- No tener que programar o utilizar comandos para comunicarse con la máquina.
- Obtención de datos de forma real y muy accesible de los seres humanos.
- Gestión del tiempo real tanto del reconocimiento como de la generación de voz.

2. Ejemplo: Asistente virtual en el hogar.



<https://www.youtube.com/watch?v=lGebo-xbzO0>

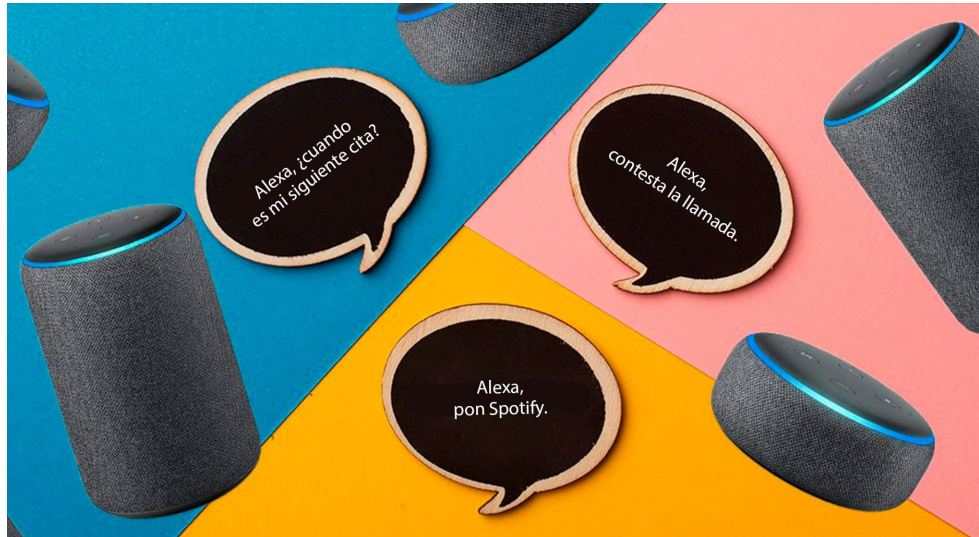
2. Ejemplo: Asistente virtual en el hogar.

Vamos a analizar cómo funcionan a nivel NLP.



2. Caso práctico: Asistente virtual en el hogar.

2.1 Activación de la comunicación: Se necesita una palabra o frase para que el dispositivo sepa que vamos iniciar una comunicación.



2. Caso práctico: Asistente virtual en el hogar.



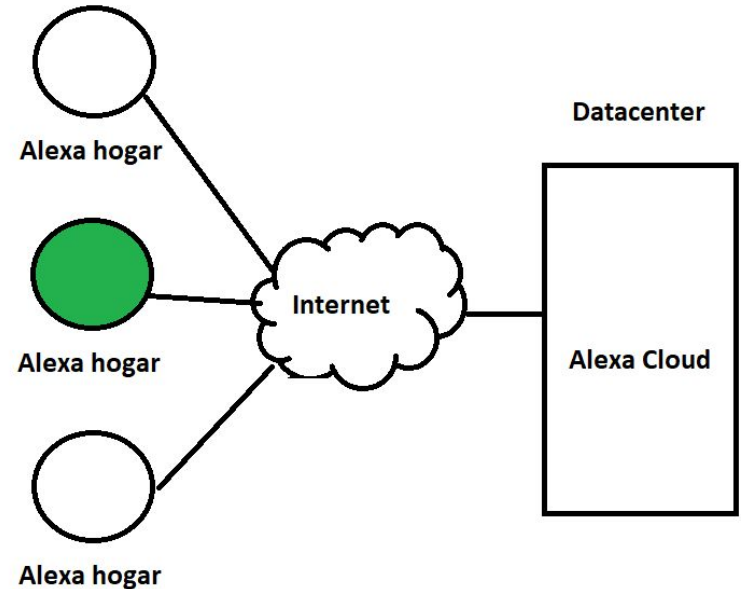
2.2 El dispositivo reconoce que se le está invocando y registra en una grabación, con las palabras de después de la palabra de activación.

- Palabra de activación: Alexa.
- Comando: ¿Qué tiempo hará mañana ?

2. Caso práctico: Asistente virtual en el hogar.

2.3 Esa grabación se envía a la nube.

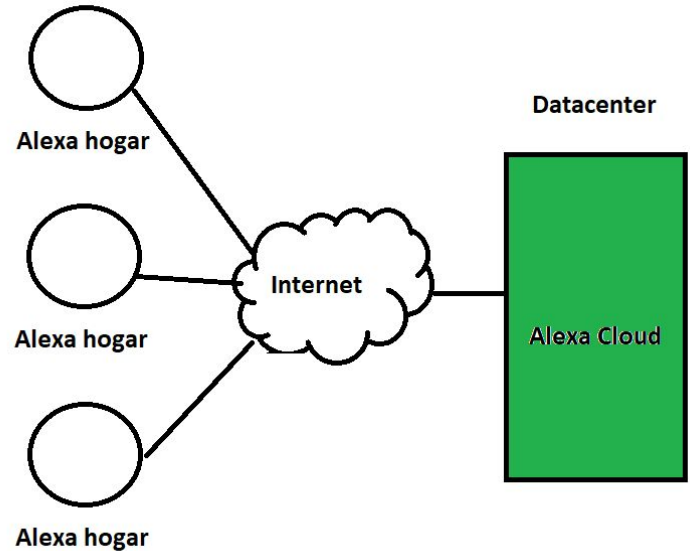
Los asistentes virtuales tienen acceso a internet ya que la mayor parte del procesamiento se realiza en servidores especializados en la nube.



2. Caso práctico: Asistente virtual en el hogar.

2.4 Procesamiento en la nube:

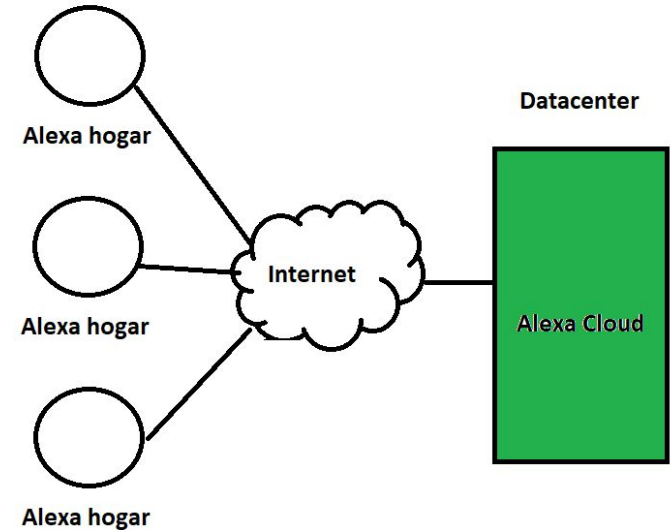
En los servidores se desglosa la frase en sonidos individuales, después se conecta a una base de datos que contiene las pronunciaciones de varias palabras para encontrar qué palabras corresponden a la combinación de sonidos individuales.



2. Caso práctico: Asistente virtual en el hogar.

2.5 Identificación palabras claves:

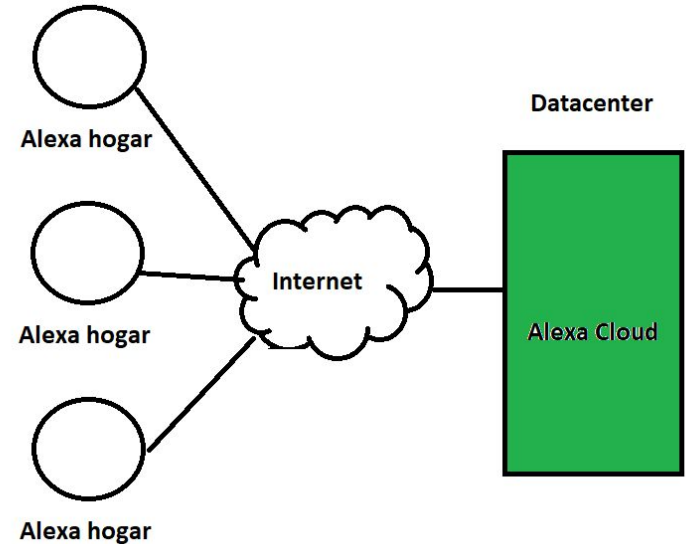
El siguiente paso es identificar las palabras más importantes para dar sentido a la tarea y poderla llevar a cabo.



2. Caso práctico: Asistente virtual en el hogar.

2.6 Resolución comando y envío al dispositivo:

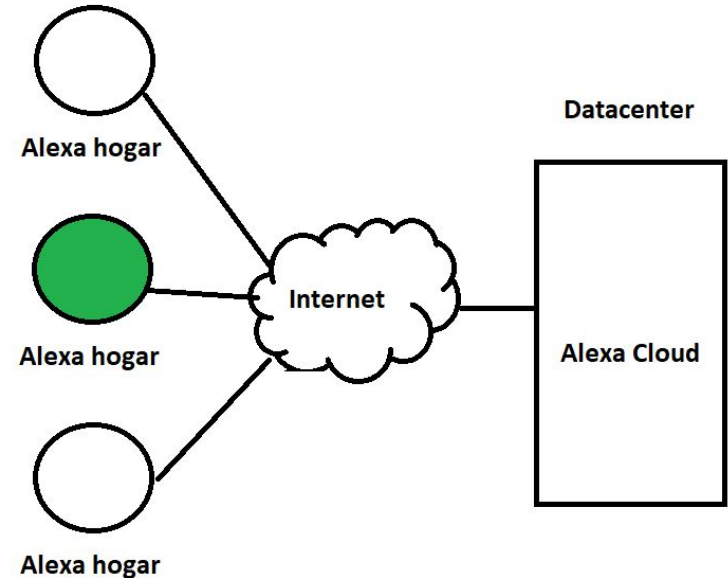
Una vez el servidor en la nube sabe la tarea a realizar y cómo llevarla a cabo, se envía la información al dispositivo del hogar.



2. Caso práctico: Asistente virtual en el hogar.

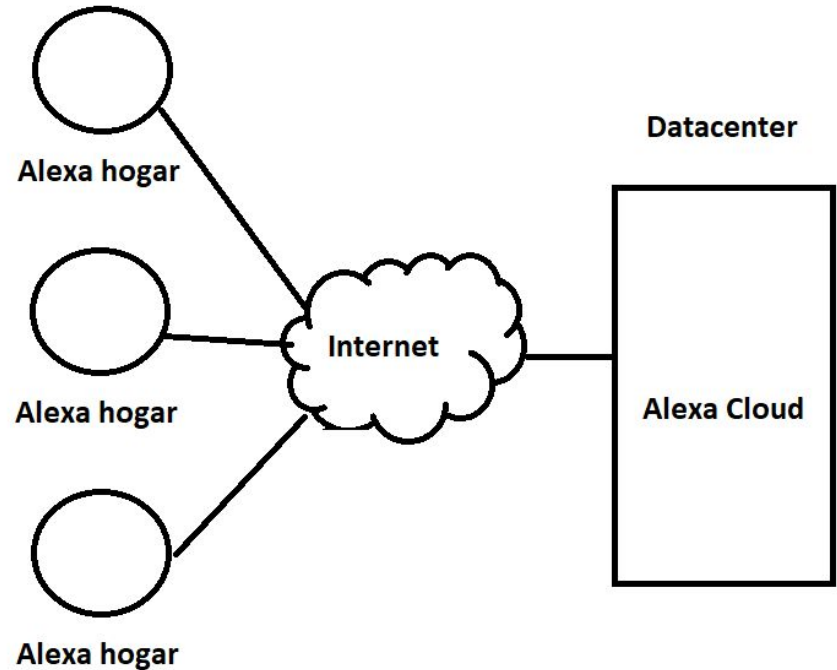
2.7 Comunicación al usuario:

El dispositivo del hogar nos comunica el resultado de la tarea.



2. Caso práctico: Asistente virtual en el hogar.

La Comprensión del lenguaje natural (ASR) y generación de texto a voz (TTS) se realiza un 95 % (aprox) en la nube de Alexa, en el dispositivo casos sencillos como la palabra de activación u órdenes de dispositivos del hogar, como por ejemplo encendido de luces.



2. Caso práctico: Asistente virtual en el hogar.

**¿ QUÉ PASA SI
SE CAE
INTERNET ?**



2. Caso práctico: Asistente virtual en el hogar.

**¿ QUÉ PASA
SI SE CAE LA
WIFI?**

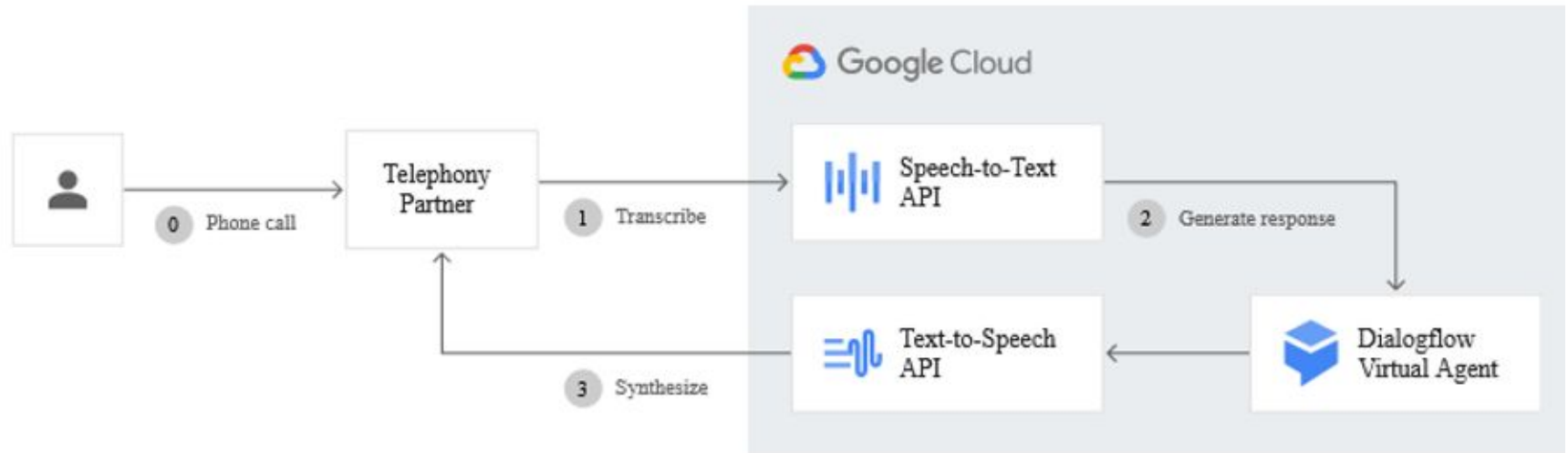


2. Caso práctico: Asistente virtual en el hogar.

**¿ CÓMO PENSÁIS
QUE FUNCIONA
OK GOOGLE?**



2. Caso práctico: Asistente virtual en el hogar.



3. Desafíos en el NLP



Como hemos visto en el ejemplo anterior, los asistentes virtuales en el entorno del hogar suponen un gran desafío en el campo de la IA del procesamiento del lenguaje natural, ya que requiere poder **comprender** al usuario y **poder ejecutar una tarea en tiempo real**.

3. Desafíos en el NLP



Existen palabras donde el significado depende de palabras anteriores o posteriores, por ejemplo “taco”, no es lo mismo decir se me ha roto el taco del zapato que decir hoy me apetece comer un taco. Otro caso especial sería las palabras “vaya”, “baya” y “valla”, la primera sería verbo “ir” la segunda es una fruta y la última un obstáculo.

3. Desafíos en el NLP



Por otro lado, existen significados de palabras que sólo existen en una región o país, un ejemplo sería la frase ¿Le provoca un tinto?

En Colombia hace referencia a tomarse un café, mientras que en otros países como España el interlocutor pensará que hace referencia a tomarse un vino tinto.

3. Desafíos en el NLP



Estos casos son innumerables en la comunicación entre seres humanos:

- Neologismos: Clickear, influencer, instagramer, castear...
- Ironías: ¡Menos mal que seguí tus consejos!, cuando realmente no ha funcionado el consejo.
- Poesía:

¿Qué es poesía?, dices mientras clavas
en mi pupila tu pupila azul.

¿Qué es poesía? ¿Y tú me lo preguntas?

Poesía... eres tú

¿ Un ser humano puede ser poesía ?

3. Desafíos en el NLP



Para solucionar estos casos del lenguaje los algoritmos en tiempo real tienen que:

- Analizar la estructura del texto para poder dividir el texto en palabras, frases y oraciones.
- Convertir ese texto en datos para poder etiquetar en categorías gramaticales.
- Encontrar el contexto de la sentencia de todos los posibles.
- Obtener la aplicación de la gramática que ha realizado el usuario, cabe resaltar que, aunque la gramática define una estructura fija su aplicación es casi infinita.

3. Desafíos en el NLP



Hay que recalcar que todas estas tareas se realizan en **tiempo real** y que cada idioma tiene su estructura de texto o categorías gramaticales por lo que el desafío es mayor.

4. Laboratorio: Ejemplos prácticos



4.1 Abrimos el buscador Google y realizamos las siguientes búsquedas:

- ¿ Qué tiempo hace en Valencia ?
- ¿Tiempo en Valencia ?
- ¿ Qué tiempo hace en Valencia Venezuela ?

¿ Google es un sistema de procesamiento del lenguaje natural ?

¿ Qué pasaría si estuviéramos en Valencia Venezuela ?

4. Laboratorio: Ejemplos prácticos



4.2 Abrimos el traductor de Google. Y realizamos la siguiente traducción.

Yo persigo una forma que no encuentra mi estilo,
botón de pensamiento que busca ser la rosa;
se anuncia con un beso que en mis labios se posa
al abrazo imposible de la Venus de Milo

I chase a way that doesn't find my style
thought button that seeks to be the rose;
is announced with a kiss that rests on my lips
to the impossible embrace of the Venus de Milo

4. Laboratorio: Ejemplos prácticos



- El poema en castellano es un cuarteto, estructura a-b-b-a, ¿ En inglés se respecta la misma estructura ?
- ¿ Las gramáticas son las misma en inglés y castellano ?
- ¿ El traductor de Google es un sistema de procesamiento del lenguaje natural ?

4. Laboratorio: Ejemplos prácticos

4.3 Ejemplo de TTS: <https://cloud.google.com/text-to-speech>

The screenshot shows the Google Cloud Text-to-Speech product page. At the top, the Google Cloud logo is on the left, and navigation links for 'Por Qué Elegir Google', 'Soluciones', 'Productos', 'Precios', and 'Primeros Pasos' are in the center. On the right, there are search, help, and language (Español) dropdowns, along with a 'Consola' button. Below the navigation bar, the page title 'Cloud Text-to-Speech' is followed by 'Contactar' and 'Empezar gratis' buttons. A yellow banner below the title reads: 'Descubre de forma exclusiva las últimas funciones de Text-to-Speech en acción en Next '21. Registrarme gratis'. The main content area is divided into three columns. The left column contains a sidebar with links: 'Text-to-Speech', 'Ventajas', 'Demostración', 'Características principales', 'Novedades', 'Documentación', 'Usos', 'Bots de voz en centros de contacto', 'Generación de voz en dispositivos', 'Guías electrónicas de programas (EPGs) accesibles', 'Todas las características', 'Precios', and 'Ve un paso más allá'. The middle column is titled 'Text-to-Speech' and describes the service: 'Convierte texto en voz que suena natural con una API basada en las tecnologías de IA de Google'. It includes two buttons: 'Probar gratis' and 'Contactar con Ventas'. Below this, there are three green checkmarks with descriptions: 'Mejora las interacciones con los clientes gracias a respuestas naturales e inteligentes.', 'Capta el interés de los usuarios con una interfaz de voz en tus dispositivos y aplicaciones.', and 'Personaliza tus comunicaciones en función de las preferencias de voz e idioma del usuario.' The right column features a Gartner logo and a quote: 'Google Cloud ha recibido la designación de líder en el Cuadrante Mágico del 2020 en el ámbito de los servicios para desarrolladores de inteligencia artificial en la nube', followed by a 'Más información' link. At the bottom, a 'VENTAJAS' section highlights three benefits: 'Voz de alta fidelidad', 'Amplia selección de voces', and 'Una voz inconfundible'.

Google Cloud

Por Qué Elegir Google Soluciones Productos Precios Primeros Pasos

Cloud Text-to-Speech

Descubre de forma exclusiva las últimas funciones de Text-to-Speech en acción en Next '21. [Registrarme gratis](#)

Text-to-Speech

Ventajas

Demostración

Características principales

Novedades

Documentación

Usos

Bots de voz en centros de contacto

Generación de voz en dispositivos

Guías electrónicas de programas (EPGs) accesibles

Todas las características

Precios

Ve un paso más allá

Text-to-Speech

Convierte texto en voz que suena natural con una API basada en las tecnologías de IA de Google

[Probar gratis](#) [Contactar con Ventas](#)

- ✓ Mejora las interacciones con los clientes gracias a respuestas naturales e inteligentes.
- ✓ Capta el interés de los usuarios con una interfaz de voz en tus dispositivos y aplicaciones.
- ✓ Personaliza tus comunicaciones en función de las preferencias de voz e idioma del usuario.

VENTAJAS

Voz de alta fidelidad

Amplia selección de voces

Una voz inconfundible

Gartner

Google Cloud ha recibido la designación de líder en el Cuadrante Mágico del 2020 en el ámbito de los servicios para desarrolladores de inteligencia artificial en la nube

[Más información](#)

4. Laboratorio: Ejemplos prácticos

4.3 Nos vamos a demostración de producto

DEMOSTRACIÓN

Prueba Text-to-Speech

Escribe lo que quieras, selecciona un idioma y haz clic en la opción de conversión a voz para oír lo que has escrito.

Text to speak:

Google Cloud Text-to-Speech enables developers to synthesize natural-sounding speech with 100+ voices, available in multiple languages and variants. It applies DeepMind's groundbreaking research in WaveNet and Google's powerful neural networks to deliver the highest fidelity possible. As an easy-to-use API, you can create lifelike interactions with your users, across many applications and devices.

text [ssml](#)

Language / locale

English (United States)

Voice type

WaveNet

Voice name

en-US-Wavenet-D

Audio device profile

Default

Speed:

1.00

Pitch:

0.00

[Show JSON](#) ▾

► SPEAK IT

4. Laboratorio: Ejemplos prácticos

4.3 Podemos los siguientes datos y damos a SPEAK IT:

Text to speak:

Google Cloud Text-to-Speech enables developers to synthesize natural-sounding speech with 100+ voices, available in multiple languages and variants. It applies DeepMind's groundbreaking research in WaveNet and Google's powerful neural networks to deliver the highest fidelity possible. As an easy-to-use API, you can create lifelike interactions with your users, across many applications and devices.

text [ssml](#)

Language / locale

English (United States)

Voice type

WaveNet

Voice name

en-US-Wavenet-D

Audio device profile

Default

Speed:

1.00

Pitch:

0.00

Show JSON ▾

► SPEAK IT

4. Laboratorio: Ejemplos prácticos

4.3 Ahora damos a SSML y vemos el código.

Text to speak:

Google Cloud Text-to-Speech enables developers to synthesize natural-sounding speech with 100+ voices, available in multiple languages and variants. It applies DeepMind's groundbreaking research in WaveNet and Google's powerful neural networks to deliver the highest fidelity possible. As an easy-to-use API, you can create lifelike interactions with your users, across many applications and devices.

text [ssml](#)

Language / locale

English (United States)

Voice type

WaveNet

Voice name

en-US-Wavenet-D

Audio device profile

Default

Speed:

1.00

Pitch:

0.00

[Show JSON](#) ▼

▶ SPEAK IT

4. Laboratorio: Ejemplos prácticos

4.3 SSML: Un estándar W3C para la generación de audios (Speech Synthesis Markup Language)

Text to speak:

```
<speech>
  <emphasis level="strong">To be</emphasis>
  <break time="200ms"/> or not to be, <break time="400ms"/>
  <emphasis level="moderate">that</emphasis>
  is the question|. <break time="400ms"/>
  Whether 'tis nobler in the mind to suffer
  The slings and arrows of outrageous fortune, <break time="200ms"/>
  Or to take arms against a sea of troubles
  And by opposing end them.
</speech>
```

text ssml

4. Laboratorio: Ejemplos prácticos

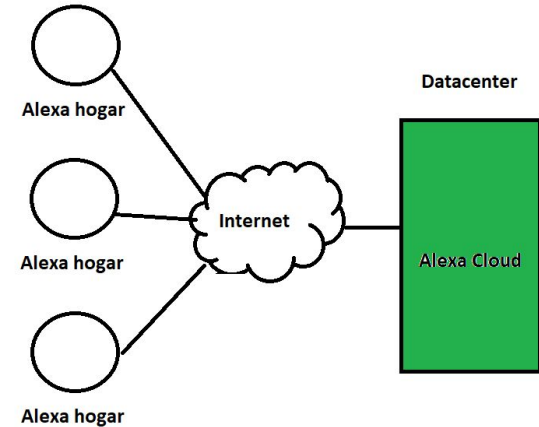
4.3 Ahora pinchamos sobre “show JSON”

```
Request URL
https://texttospeech.googleapis.com/v1beta1/text:synthesize

Request body
{
  "audioConfig": {
    "audioEncoding": "LINEAR16",
    "pitch": 0,
    "speakingRate": 1
  },
  "input": {
    "ssml": "<speak> <emphasis level='strong'>To be</emphasis> <break time='200ms'>/> or not to be, <break time='400ms'>/> <emphasis level='moderate'>that</emphasis> is the question.<break time='400ms'>/> Whether 'tis nobler in the mind to suffer The slings and arrows of outrageous fortune,<break time='200ms'>/> Or to take arms against a sea of troubles And by opposing end them. </speak>"
  },
  "voice": {
    "languageCode": "es-ES",
    "name": "es-ES-Standard-C"
  }
}
```

4. Laboratorio: Ejemplos prácticos

4.3 JSON (JavaScript Object Notation) es un formato ligero de intercambio de datos. Gracias a esta tecnología podemos intercambiar datos mediante API. Así se envía la información entre el hogar y el cloud en el ejemplo visto en el caso práctico.



4. Laboratorio: Ejemplos prácticos

4.3 Por último haz pruebas con los siguientes parámetros expuestos en la imagen.

Text to speak:

```
<say>  
  <emphasis level="strong">To be</emphasis>  
  <break time="200ms"/> or not to be, <break time="400ms"/>  
  <emphasis level="moderate">that</emphasis>  
  is the question.<br>  
  Whether 'tis nobler in the mind to suffer  
  The slings and arrows of outrageous fortune,<br>  
  Or to take arms against a sea of troubles  
  And by opposing end them.  
</say>
```

text ssm1

Language / locale
Español (España)

Voice type
Basic

Voice name
es-ES-Standard-C

Audio device profile
Default

Speed: 1.00 Pitch: 0.00

Show JSON

► SPEAK IT

4. Laboratorio: Ejemplos prácticos



¿ Cambia la pronunciación ?

¿ Cuando cambiamos un parámetro cambia el JSON ?

¿ Y el contenido del SSML ?

4. Laboratorio: Ejemplos prácticos

4.4 Ejemplo de ASR: <https://cloud.google.com/speech-to-text>

The screenshot shows the Google Cloud website for the Speech-to-Text service. The header includes the Google Cloud logo, navigation links (Por Qué Elegir Google, Soluciones, Productos, Precios, Primeros Pasos), a search icon, and a language dropdown set to 'Español'. The main content area is titled 'Speech-to-Text' and describes the service as converting voice to text with a precise API. It features a 'Probar gratis' button and a 'Contactar con Ventas' button. A sidebar on the left lists various sections: Ventajas, Demostración, Características principales, Documentación, Usos, and Todas las características. A Gartner badge on the right mentions the 2021 Magic Quadrant. At the bottom, three key benefits are highlighted: 'Precisión de última generación', 'Personalización de modelos sencilla', and 'Despliegue flexible'.

Google Cloud

Por Qué Elegir Google Soluciones Productos Precios Primeros Pasos

Cloud Speech-to-Text

Speech-to-Text

Convierte voz en texto de forma precisa con una API basada en las tecnologías de IA de Google

Probar gratis Contactar con Ventas

✓ Transcribe tu contenido y ofrece subtítulos precisos.

✓ Mejora la experiencia de tus usuarios con comandos de voz.

✓ Obtén información valiosa a partir de las interacciones con los clientes y mejora tu servicio

VENTAJAS

Precisión de última generación

Speech-to-Text aplica los algoritmos de la red neuronal de aprendizaje profundo más avanzada de Google para

Personalización de modelos sencilla

La UI de Speech-to-Text te permite experimentar, crear y gestionar recursos personalizados.

Despliegue flexible

Despliega reconocimiento de voz donde lo necesites, ya sea en la nube con la API u on-premise con [Speech-to-Text On-Prem](#).

Gartner

En el Cuadrante Mágico del 2021, Gartner designa a Google como uno de los líderes en servicios para desarrolladores de IA en la nube.

[Regístrate para descargar el informe](#)

Ventajas

Demostración

Características principales

Descubre cómo integran los clientes Speech-to-Text en sus soluciones de audio y voz

Novedades

Documentación

Usos

Mejora del servicio de atención al cliente

Activar control por voz

Transcripción de contenido multimedia

Todas las características

Precios

Ve un paso más allá

4. Laboratorio: Ejemplos prácticos

4.4 Nos vamos a demostración.

DEMOSTRACIÓN

Comprueba cómo funciona Speech-to-Text en la práctica

Al igual que en esta demostración, puedes integrar fácilmente la transcripción de voz en tus aplicaciones con la API Speech-to-Text.

Input type

☐ Microphone ☒ File upload

Language

English (United States) ▼

Speaker diarization BETA

Off ▼

Speakers

1 speaker ▼

Punctuation

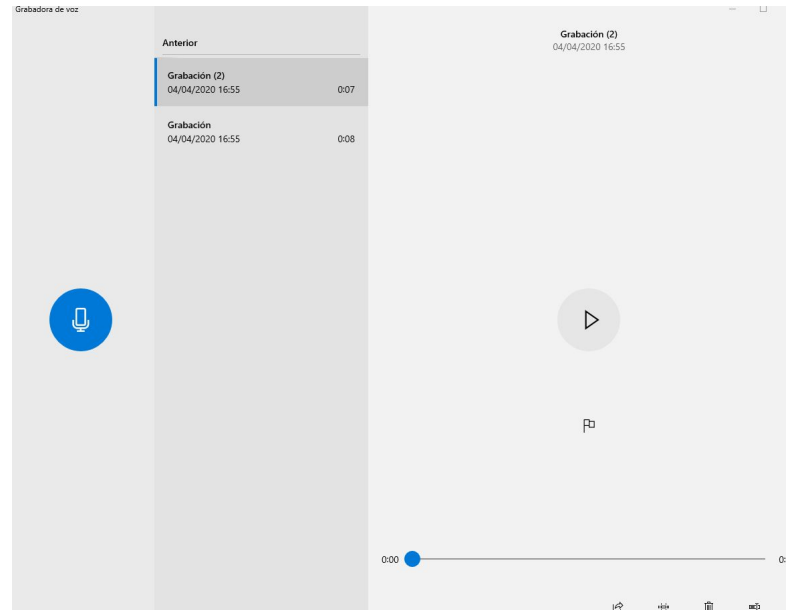
☒

Show JSON ▼

↑ CHOOSE FILE

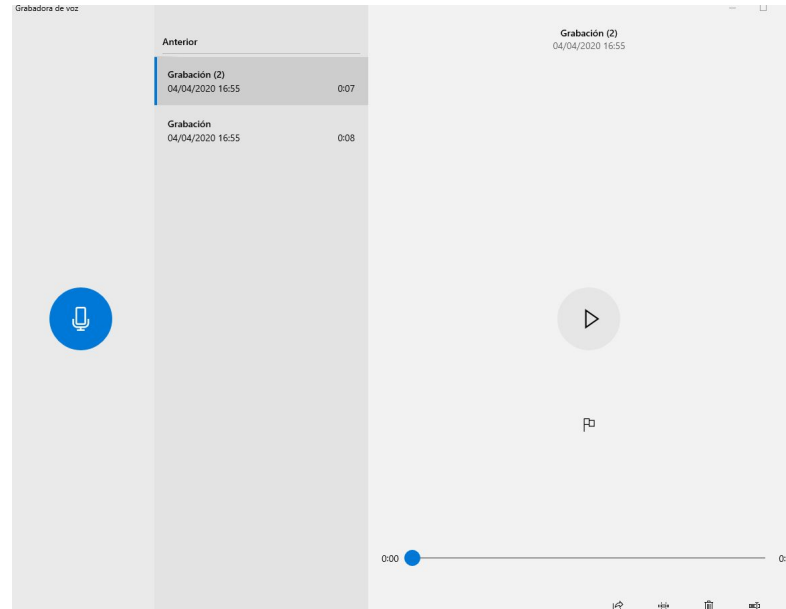
4. Laboratorio: Ejemplos prácticos

4.4 Abrimos la grabadora de Windows.



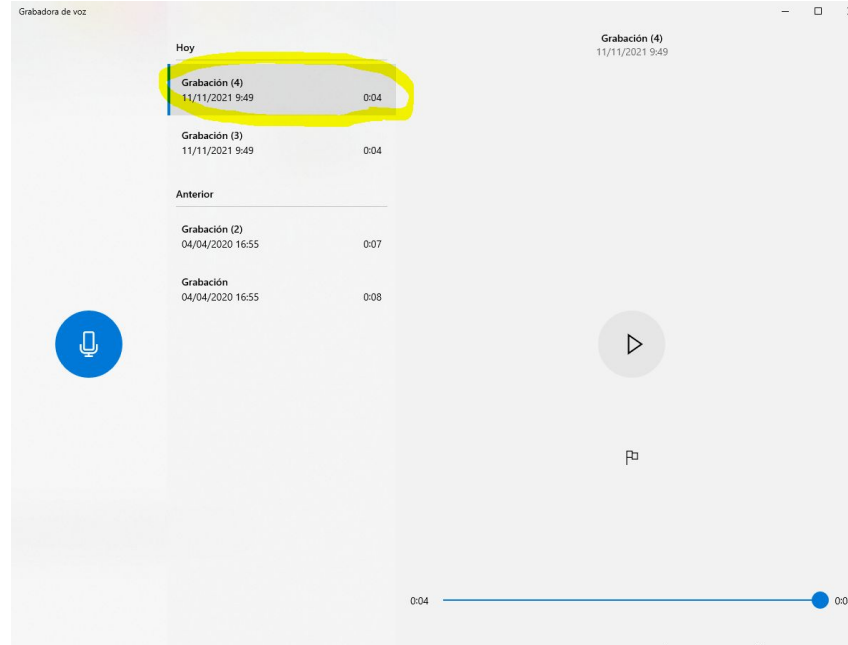
4. Laboratorio: Ejemplos prácticos

4.4 Realizamos una grabación como si fuera una nota de voz.



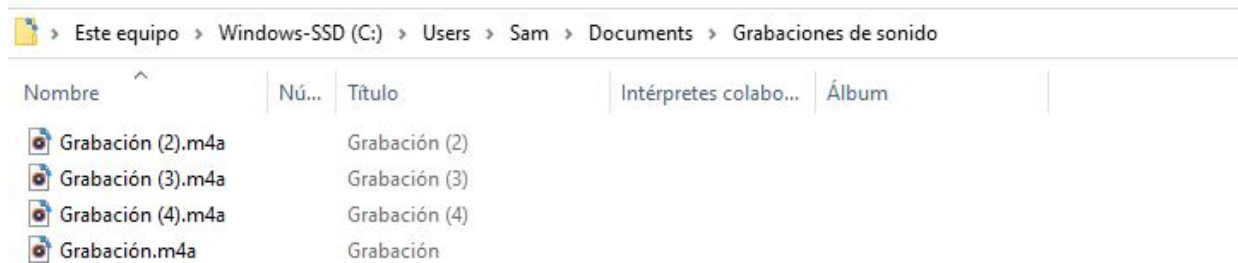
4. Laboratorio: Ejemplos prácticos

4.4 Seleccionamos la grabación y pulsamos sobre abrir ubicación en destino.



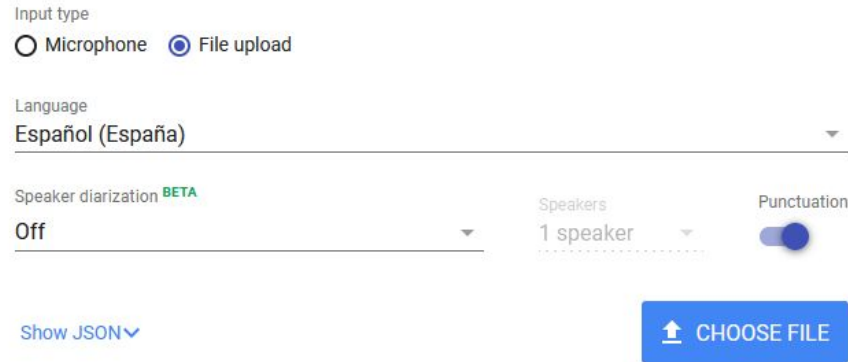
4. Laboratorio: Ejemplos prácticos

4.4 Seleccionamos la grabación y pulsamos sobre abrir ubicación en destino.



4. Laboratorio: Ejemplos prácticos

4.4 En la página de Google, seleccionamos Español (España) y sube el fichero.



Input type

☐ Microphone ☒ File upload

Language

Español (España)

Speaker diarization BETA

Off

Speakers

1 speaker

Punctuation

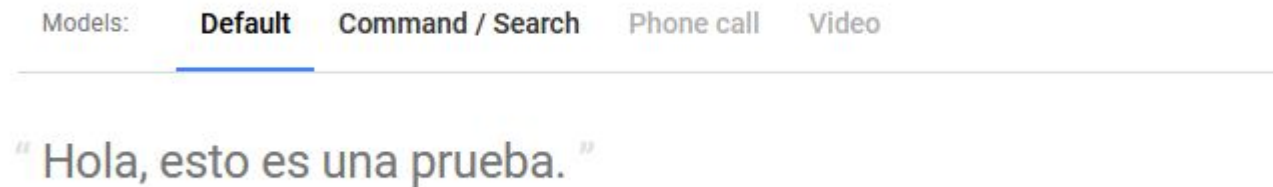
☒

Show JSON

CHOOSE FILE

4. Laboratorio: Ejemplos prácticos

4.4 ¿Cómo ha ido ?.



4. Laboratorio: Ejemplos prácticos



4.4 Con el mismo fichero si lo volvemos a procesar con el idioma en inglés ¿ Qué ocurre ?.