Aplicaciones de análisis: Computación Cognitiva

Servicios de Voz

Servicios deVoz

Voz a Texto: Convertir la voz humana en palabras escritas

Speech to Text para ser utilizado en cualquier escenario donde sea necesario tender un puente entre la palabra hablada y su forma escrita

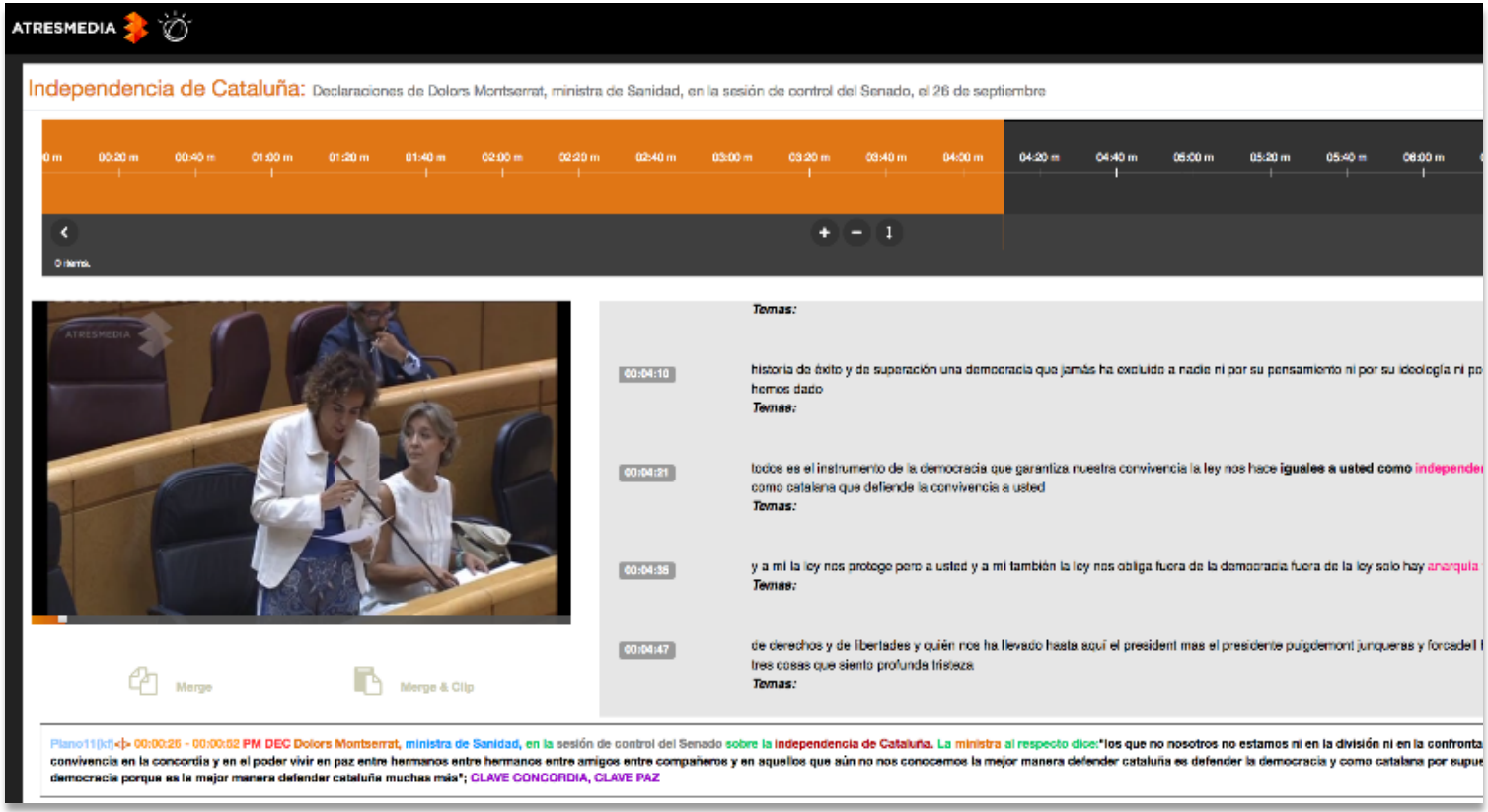
Uso Previsto	
Aprovecha la inteligencia artificial para combinar la información sobre la gramática y la estructura del lenguaje con el conocimiento de la señal de audio. El servicio devuelve continuamente y actualiza retroactivamente la transcripción a medida que se escucha más discurso. Ofrece una rica interfaz de personalización para la adaptación del dominio	Presenta tres interfaces • **WebSocket** • **HTTP REST** • **Asynchronous HTTP**

Casos de Uso más habituales	Input	Output
• Interacciones con móviles • Dotar a los chat bots con voz • **Transcribir ficheros de grabación** • **Transcripciones de Centros de Llamadas** • **Control de voz embebida en sistemas** • Convertir sonido a texto para hacerlo consultable	• Audio en stream con una velocidad inteligible • Audio registrado con una velocidad inteligible	Transcripciones de texto del audio con las palabras reconocidas

El trabajo actual del documentalista

Ratio trabajo,1 x 3 (Para generar clips de 1 hora de video bruto se tarda 3 horas de trabajo de un documentalista).

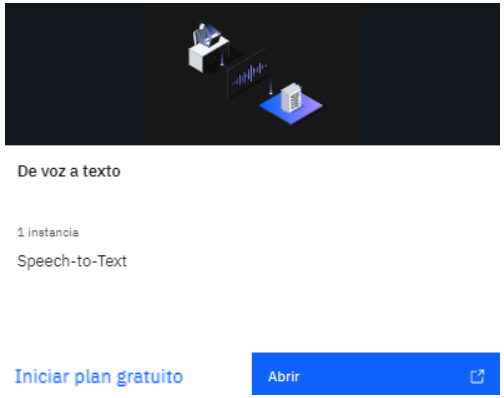
Plano 51 (kf)<|>
00:59:36.22 - 01:01:40.04 PM DEC **Dolors Montserrat, Ministra de Sanidad**, Servicios Sociales e Igualdad sobre el ACOSO SEXUAL y VIOLENCIA DE GÉNERO reprochando a Podemos que no participen en el pacto de Estado contra la violencia de género: **"Señoría, no cabe duda de que la violencia que sufrimos las mujeres va mucho más allá de la que se produce en el marco de una relación de pareja; La trata de mujeres y niñas con fines de explotación sexual, la mutilación genital femenina o las agresiones sexuales, acoso y violaciones hacen sufrir de forma terrible a mujeres de todos los rincones del mundo, pero le voy a decir una cosa, y es que esta cuestión nos afecta a todos como sociedad, por eso podríamos avanzar desde la unidad; unidad que ustedes rompieron; Esta unidad está representada de la mejor forma posible en el pacto de Estado en el que estamos trabajando todas las fuerzas políticas y todas las comunidades autónomas, el Pacto de Estado contra la violencia de género, en el que ustedes son los únicos de esta Cámara que se han abstenido y que realmente no han votado a favor; Ustedes tenían la respuesta; Usted me pregunta dónde están las respuestas; Están en el pacto de Estado, al que ustedes han dicho que no, se han abstenido; Por tanto, no nos venga a dar lecciones en esta lucha porque nosotros sí que estamos trabajando cada día en ello; Un pacto de Estado que en el ámbito de las agresiones sexuales contiene más de veinte medidas tan importantes como el establecimiento de protocolos contra el acoso sexual en las empresas, la introducción de los cambios legislativos pertinentes para la aplicación del Convenio de Estambul y la promoción de programas integrales de atención a la violencia sexual; Le voy a decir más; las agresiones sexuales, el abuso y el acoso sexual ya están castigados en nuestro ordenamiento jurídico con penas de prisión que pueden llegar hasta los diez años; Además, en la anterior legislatura se siguió ampliando esta protección, elevando el consentimiento sexual de los trece a los dieciséis años, se prohibió trabajar con niños a personas que hayan sido condenadas por delitos contra la libertad sexual y se estableció que los menores víctimas de agresiones sexuales no mantengan ningún tipo de contacto con su agresor en el momento de declarar para evitar su victimización; Por tanto, estamos avanzando también en la prevención y en la educación basada en el respeto y en la construcción de unas relaciones afectivas sanas; Espero y deseo que se sumen al pacto de Estado porque esta es la respuesta y este es el instrumento"** CLAVE CURIOSIDADES. CLAVE GESTOS. CLAVE MODA


ATRESMEDIA
Independencia de Cataluña: Declaraciones de Dolors Montserrat, ministra de Sanidad, en la sesión de control del Senado, el 26 de septiembre
Temas:
00:04:10
historia de éxito y de superación una democracia que jamás ha excluido a nadie ni por su pensamiento ni por su ideología ni po
hemos dado
Temas:
00:04:21
todos es el instrumento de la democracia que garantiza nuestra convivencia la ley nos hace iguales a usted como independer
como catalana que defiende la convivencia a usted
Temas:
00:04:35
y a mí la ley nos protege pero a usted y a mí también la ley nos obliga fuera de la democracia fuera de la ley solo hay anarquía
Temas:
00:04:47
de derechos y de libertades y quién nos ha llevado hasta aquí el president mas el presidente puigdemont junqueras y forcadell
tres cosas que siento profunda tristeza
Temas:
Merge
Merge & Clip
Plano11[kf]<|> 00:00:26 - 00:00:52 PM DEC Dolors Montserrat, ministra de Sanidad, en la sesión de control del Senado sobre la independencia de Cataluña. La ministra al respecto dice:"los que no nosotros no estamos ni en la división ni en la confronta
convivencia en la concordia y en el poder vivir en paz entre hermanos entre hermanos entre amigos entre compañeros y en aquellos que aún no nos conocemos la mejor manera defender cataluña es defender la democracia y como catalana por supue
democracia porque es la mejor manera defender cataluña muchas más"; CLAVE CONCORDIA, CLAVE PAZ

Instrucciones. Instanciación del servicio

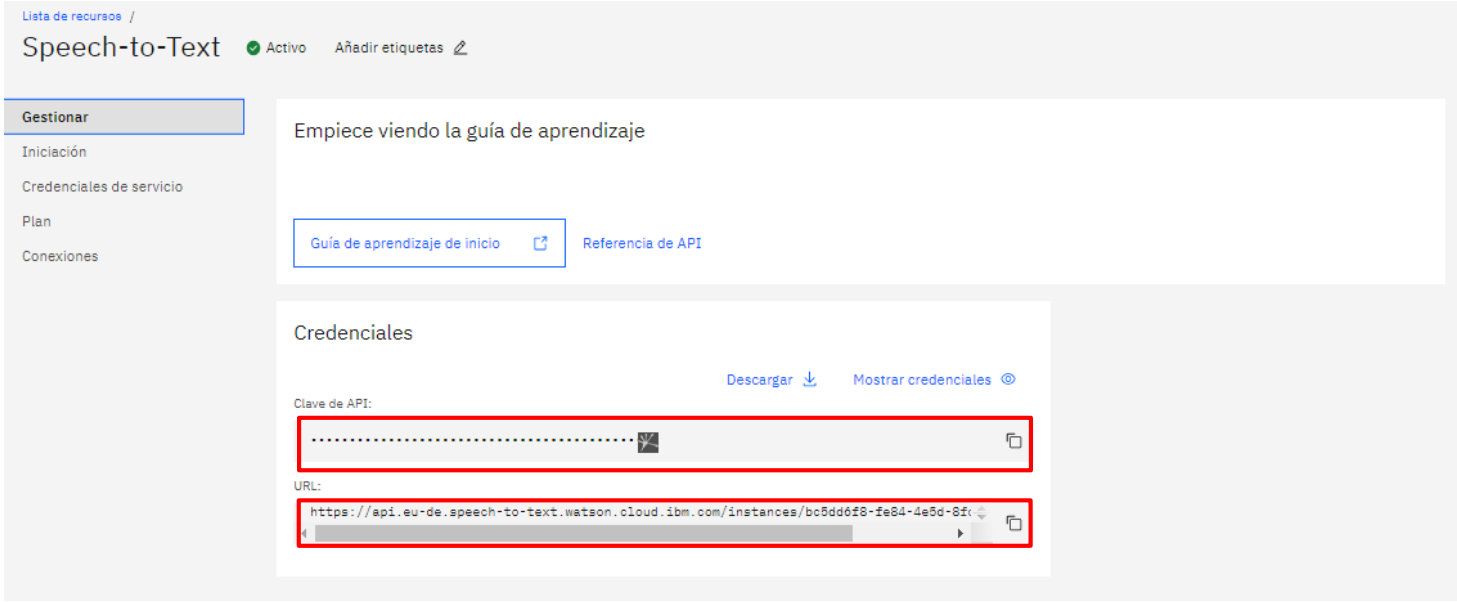
- Conectate al servicio de IBM Cloud (https://cloud.ibm.com)
- Accede a https://cloud.ibm.com/developer/watson/dashboard

  Y selecciona Iniciar plan gratuito de “De voz a texto”



Instrucciones. Toma nota del API Token y la URL

- Una vez instanciado el servicio accede a su página principal y toma nota del API Token y la url del servicio instanciado



Invocando el servicio directamente

Tenemos varias alternativas:

- Utilizar cualquiera de los SDKs disponibles
- Hacer pruebas unitarias a través de comandos o herramientas:
  - Comando curl (ver para un ejemplo):
    https://cloud.ibm.com/docs/speech-to-text?topic=speech-to-text-gettingStarted
  - Herramientas tipo Postman, por ejemplo
- Utilizar demos específicas disponibles en git hub (nodejs)
- Utilizar demos en la web
  - Todos los idiomas:
    https://speech-to-text-demo.ng.bluemix.net/
  - Solo inglés:
    https://www.ibm.com/demos/live/speech-to-text/self-service/home

Invocando el servicio desde una aplicación ejemplo

- Instalación de nodejs:
  - https://nodejs.org/es/download
- Aconsejable versión superior a v16
- Descarga el ejemplo de github. Utiliza el comando git clone
  - https://github.com/IBM/speech-to-text-code-pattern
  - https://github.com/IBM/text-to-speech-code-pattern
- Configurar los valores de usuario y password (o API key) en el fichero .env si es ejecución local

Nota: En algunos casos puede dar problemas el uso del micrófono en entornos local. En esos casos utilizar la demo disponible en internet.

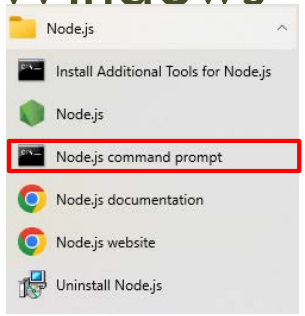
Instrucciones. Descarga e instalación de la aplicación

- En tu portátil, abre una terminal y clona la aplicación a descargar

```
git clone https://github.com/IBM/speech-to-text-code-pattern
```

- Dentro del subdirectorio creado copia el fichero .env.example como .env, descomenta las líneas de IBM Cloud y pon tus variables de API Token y URL

Instalación de la aplicación nodejs ejemplo en Windows



1. Abre una terminal especial de nodejs:
2. Situate en el directorio donde has descargado la aplicación
3. Ejecuta en la terminal:

```
npm install -g npm latest
```

4. Borra el fichero package-lock.json
5. Modifica el fichero package json (las modificaciones están en rojo)

   Cambia la sintaxis de la línea "build:" incorporando las modificaciones en rojo

   "build": "INLINE_RUNTIME_CHUNK=false react-scripts build",

   Por

   "build": "**(set** INLINE_RUNTIME_CHUNK=false**)** **&&** craco build",

6. Ejecuta, finalmente:

```
npm install
npm install craco
npm run build
npm start
```

Instalación de la aplicación nodejs ejemplo en Linux

- Sitúate en el directorio donde has descargado la aplicación
- Sigue los siguientes pasos para asegurarnos que se despliega adecuadamente la instalación en Linux:

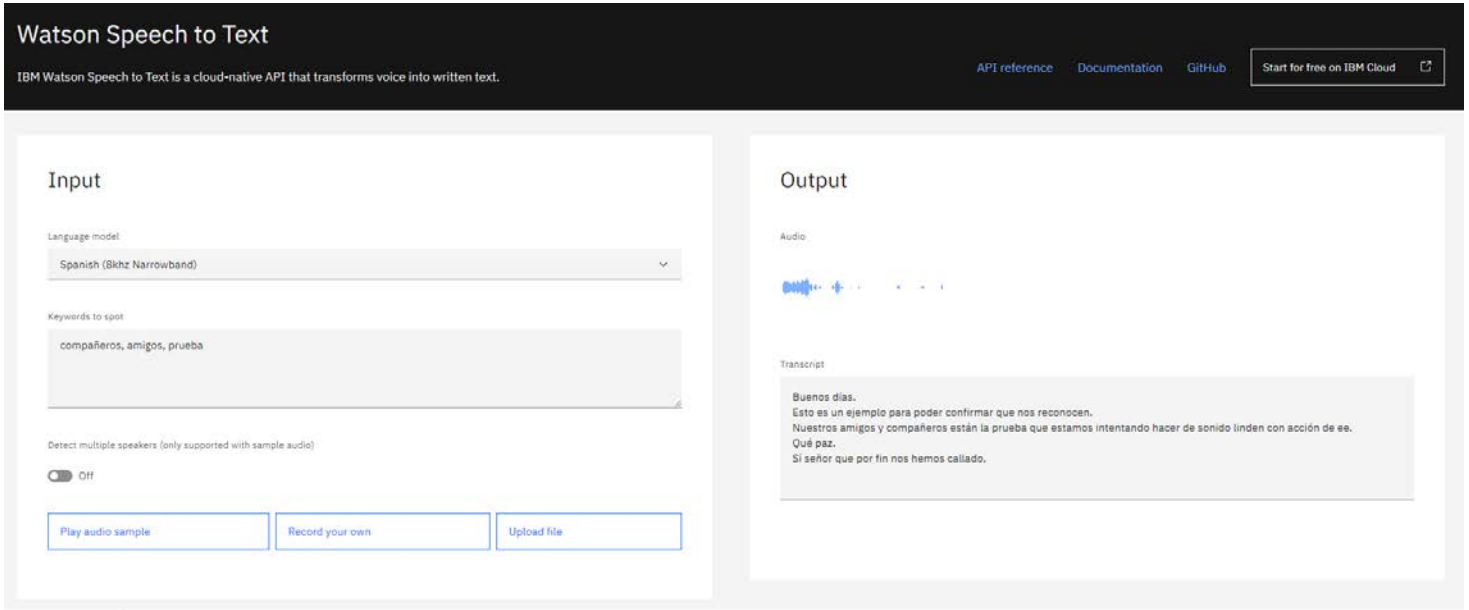
```
sudo npm install -g npm latest
```

- Borra el fichero package-lock.json
- Ejecuta, a continuación:

```
npm install
npm install craco
npm run build
npm start
```

Ejemplo StT

- Habilita en la imagen virtual el micrófono: (Dispositivos -> Audio -> Entrada de Audio)



Invocación con Curl. Ejemplos.

```
curl -X POST -u "apikey:_m_ygizLpIXx96DKZg7GqfgoJK-1Hgw4mbzYcU27QUpz" \
        --header "Content-Type: audio/flac" \
        --data-binary @audio-file.flac \
        https://api.eu-gb.speech-to-text.watson.cloud.ibm.com/instances/899315cf-8053-4f99-8e7c-2c89bb3c11f5/v1/recognize
```

```
curl -X POST -u "apikey:_m_ygizLpIXx96DKZg7GqfgoJK-1Hgw4mbzYcU27QUpz" \
        --header "Content-Type: audio/mp3" \
        --data-binary @Prueba_voz.mp3 \
        --output "Prueba_voz.json" \
        https://api.eu-gb.speech-to-text.watson.cloud.ibm.com/instances/899315cf-8053-4f99-8e7c-2c89bb3c11f5/v1/recognize?model=es-ES_NarrowbandModel
```

Opciones de entrenamiento de StT

- Entrenamiento del Modelo de Lenguaje

  Poder crear un nuevo modelo de lenguaje basado en un idioma determinado y al que se pueden incorporar modificaciones al modelo de lenguaje base. Indicado para incluir nuevos juegos de palabras específicos para un dominio en concreto (Salud, Industria)

- Entrenamiento del Modelo Acústico

  Para adaptarlo al entorno acústico y al acento del hablante

- Entrenamiento gramatical

  Para adaptarlo a respuestas cortas correspondientes a un modelo predeterminado en la mayoría de las ocasiones

Frecuencia del sonido

- Teorema de muestreo de Nyquist-Shannon: para poder reproducir una señal periódica ésta deberá ser limitada en frecuencia y la tasa de muestreo deberá ser el doble de su ancho de banda
- La frecuencia de la grabación es clave, por tanto, para poder realizar la transcripción:
  - 8 KHz / 16 KHz. Frecuencia habitual en el entorno conversacional (telefónico)
  - 44,1KHz. Frecuencia más adecuada para entornos musicales
- Muchas de las soluciones comerciales de grabación de sonido establecen las condiciones de grabación y reproducción: frecuencia, bitrate y formato (especialmente en soluciones multimedia)

Calidad de la transcripción

- Es habitual tener que evaluar la calidad de los servicios de reconocimiento de voz automática (ASR) en nuestro proyecto debido a la influencia del vocabulario (jerga), entorno acústico, pronunciación, etc.
- Uno de los métodos para poder evaluar la precisión de la transcripción (exclusivamente) es a través del WER (Word Error Rate).
- ¿Qué podemos esperar de la comparación de la calidad?:
  - La transcripción humana ronda una tasa de error del 4%
  - Rangos inferiores al 5-4% se pueden conseguir con entornos específicos de lenguaje y con tasas muy importantes de entrenamiento
  - Es fácil conseguir un promedio de error de un 10-20% en los primeros intentos

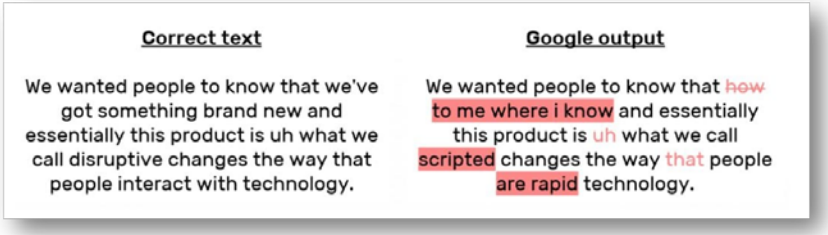
WER: Word Error Rate

- La evaluación de cualquier ASR tiene que ser, en general:
  - Directa: Independiente de la aplicación ASR
  - Objetiva: Se puede hacer de forma automatizada
  - Interpretable: Es un valor numérico
  - Modular
- Se basa en palabras (no en fonemas)
- Distancia de Levenshtein: cuantas modificaciones hay que hacer a una palabra para llegar a convertirse en otra. La distancia de Levensthein entre "casa" y "calle" es de 3:
  - casa → cala (**sustitución** de 's' por 'l')
  - cala → calla (**inserción** de 'l' entre 'l' y 'a')
  - calla → calle (**sustitución** de 'a' por 'e')

WER: Word Error Rate

- Las operaciones reconocidas a nivel de palabras son:
  - Sustitución
  - Borrado
  - Inserción

WER = (S + B + I) / Numero de palabras



Correct text	Google output
We wanted people to know that we've got something brand new and essentially this product is uh what we call disruptive changes the way that people interact with technology.	We wanted people to know that ~~how~~ to me where i know and essentially this product is uh what we call scripted changes the way that people are rapid technology.

- WER:
  - Palabras en total: 29
  - WER: (6+2+3)/29

"we've got something brand new" - "how to me where I know": 5-S + 1-I

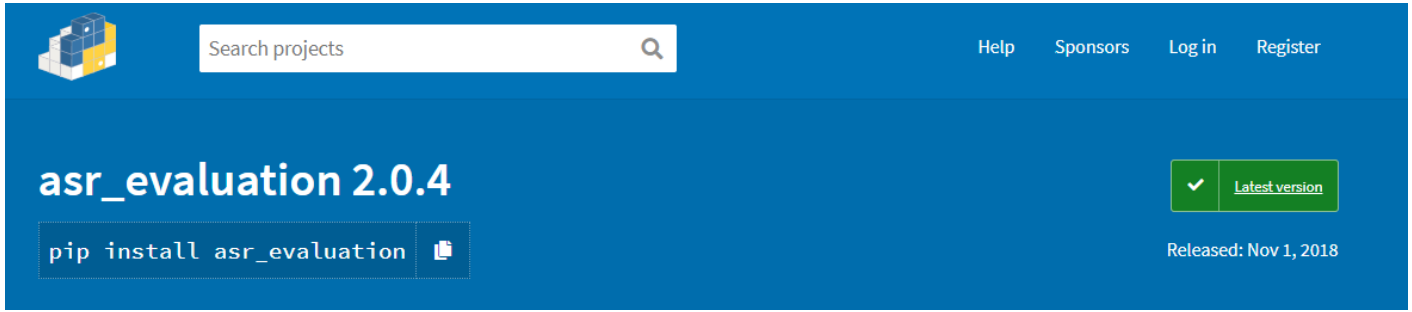
"disruptive" - "scripted": 1-B + 1-I

"interact" - "are rapid": 1-B + 2-I

Qué no se está considerando en el WER

- La calidad de la grabación
- La sensibilidad del micrófono
- Pronunciación del interlocutor
- Ruido o sonido de fondo
- Nombres, localizaciones y nombres propios no usuales
- Términos específicos de la industria o técnicos

Ejemplos de implementación



- The program outputs three standard measurements:
  - Word error rate (WER)
  - Word recognition rate (the number of matched words in the alignment divided by the number of words in the reference).
  - Sentence error rate (SER) (the number of incorrect sentences divided by the total number of sentences).

Servicios deVoz

Texto a Voz: Habilitar a los sistemas con voz humana

Text to Speech convertir el texto escrito en voz natural con una variedad de lenguajes y voces

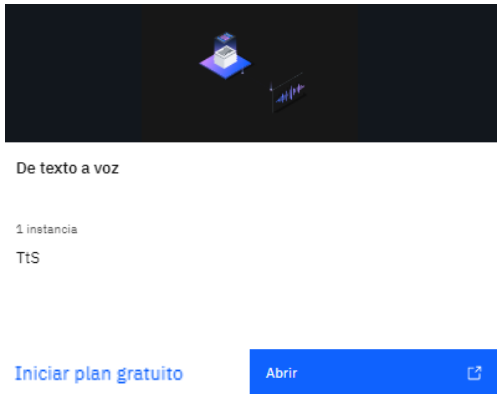
Uso Previsto	
El servicio proporciona una API que utiliza las capacidades de síntesis de voz de IBM para convertir el texto escrito en voz natural. El servicio transmite los resultados al cliente con un retraso mínimo. Una interfaz de personalización le permite especificar cómo se pronuncian las palabras inusuales que aparecen en su entrada, así como los estilos expresivos	Dos interfaces • **WebSocket** • **HTTP REST**

Casos de Uso más habituales	Input	Output
• Herramientas de ayuda para personas con problemas de vision • **Dotar de voz a los chat bots** • **Crear herramientas educativas basadas en la lectura** • **Aplicaciones en móviles** • **Comunicar direcciones o información con las manos libres** • Desarrollar juguetes interactivos para niños	Cualquier texto	Voz con diferentes alternativas de voces

Instrucciones. Instanciación del servicio

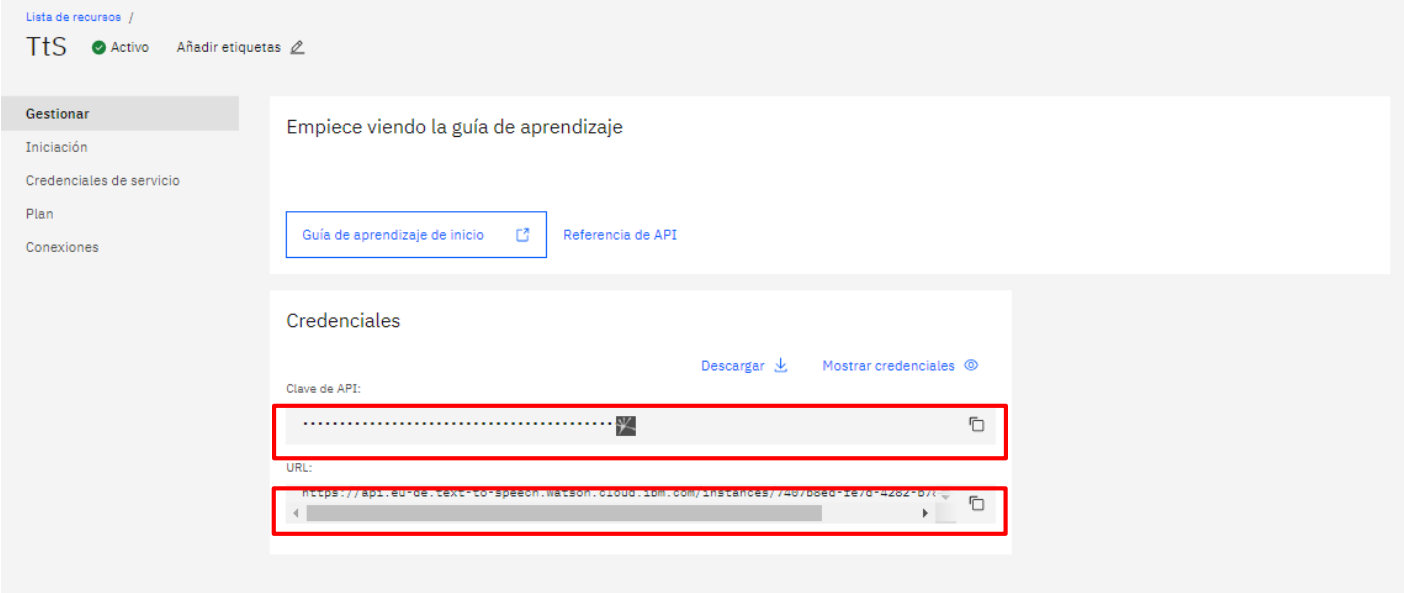
- Conectate al servicio de IBM Cloud (https://cloud.ibm.com)
- Accede a https://cloud.ibm.com/developer/watson/dashboard

  Y selecciona Iniciar plan gratuito de “De texto a voz”



Instrucciones. Toma nota del API Token y la URL

- Una vez instanciado el servicio accede a su página principal y toma nota del API Token y la url del servicio instanciado



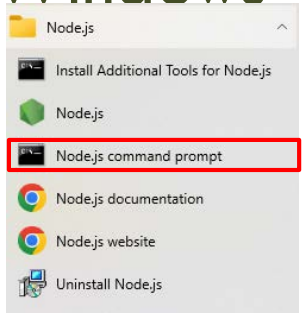
Instrucciones. Descarga e instalación de la aplicación

- En tu portátil, abre una terminal y clona la aplicación a descargar

```
git clone https://github.com/IBM/speech-to-text-code-pattern
```

- Dentro del subdirectorio creado copia el fichero .env.example como .env, descomenta las líneas de IBM Cloud y pon tus variables de API Token y URL

Instalación de la aplicación nodejs ejemplo en Windows



1. Abre una terminal especial de nodejs:
2. Situate en el directorio donde has descargado la aplicación
3. Ejecuta en la terminal:

```
npm install -g npm latest
```

4. Borra el fichero package-lock.json
5. Modifica el fichero package json (las modificaciones están en rojo)

   Cambia la sintaxis de la línea "build:" incorporando las modificaciones en rojo

   "build": "INLINE_RUNTIME_CHUNK=false react-scripts build",

   Por

   "build": "**(set** INLINE_RUNTIME_CHUNK=false**)** **&&** react-scripts build",

6. Ejecuta, finalmente:

```
npm install
npm install react-scripts
npm run build
npm start
```

Instalación de la aplicación nodejs ejemplo en Linux

- Sitúate en el directorio donde has descargado la aplicación
- Sigue los siguientes pasos para asegurarnos que se despliega adecuadamente la instalación en Linux:

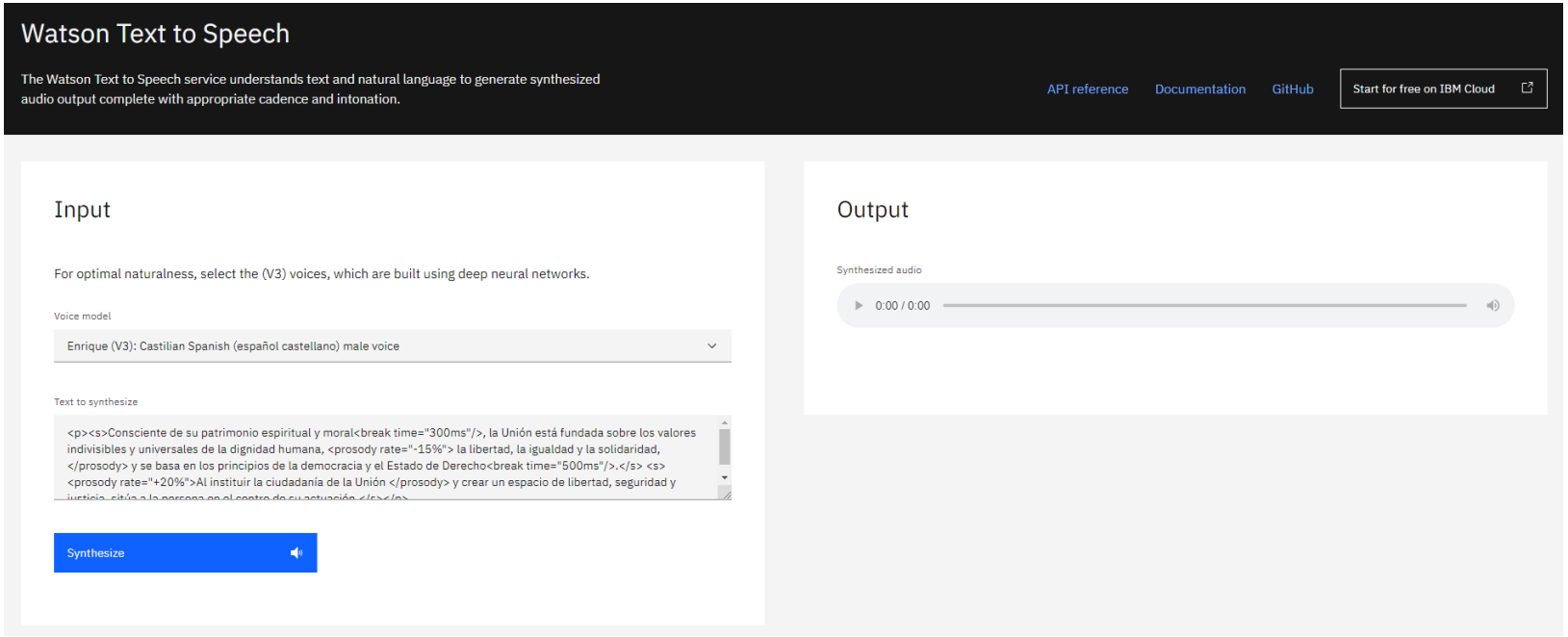
```
sudo npm install -g npm latest
```

- Borra el fichero package-lock.json
- Ejecuta, a continuación:

```
npm install
npm install craco
npm run build
npm start
```

Ejemplo de TtS

- Utiliza el texto libre y añade tags de SSML para comprobar el efecto



Invocación con Curl

```
curl -X POST -u "apikey:FyKXNEmsqY6p5hHzkQLOhUdPkJ89FRXOzyC217ikdEky" \
        --header "Content-Type: application/json" \
        --header "Accept: audio/wav" \
        --data "{\"text\":\"hello world\"}" \
        --output hello_world.wav \
        https://api.eu-gb.text-to-speech.watson.cloud.ibm.com/instances/0642d19b-5f35-49aa-a53c-34fb024a4331/v1/synthesize
```

```
curl -X POST -u "apikey:FyKXNEmsqY6p5hHzkQLOhUdPkJ89FRXOzyC217ikdEky" \
        --header "Content-Type: application/json" \
        --header "Accept: audio/wav" \
        --data "{\"text\":\"Bienvenido al ejemplo de oir hablar al ordenador\"}" \
        --output Bienvenido_spanish_Enrique.wav \
        https://api.eu-gb.text-to-speech.watson.cloud.ibm.com/instances/0642d19b-5f35-49aa-a53c-34fb024a4331/v1/synthesize?voice=es-ES_EnriqueV3Voice
```

Dando más expresividad

- SSML (Speech Synthesis Markup Language) es un lenguaje basado en XML que proporciona anotaciones de texto para aplicaciones de síntesis de voz.
- Es una recomendación del grupo de trabajo del W3C Voice-Browser adoptada como lenguaje de marcado estándar para la síntesis del habla por la especificación VoiceXML 2.0.
- Por ejemplo, se puede añadir más expresividad con tags del tipo:

```
<express-as type="GoodNews">
  I am pleased to inform you that your mortgage loan application was approved.
</express-as>
```

- Tambien hay otras opciones, como Apology, Uncertainty…

Invocación con Curl

curl -X POST -u "apikey:**FyKXNEmsqY6p5hHzkQLOhUdPkJ89FRXOzyC217ikdEky**" \

--header "Content-Type: application/json" \

--header "Accept: audio/wav" \

--data **"@SSML.json"** \

--output Spanish_SSML.wav \

https://api.eu-gb.text-to-speech.watson.cloud.ibm.com/instances/0642d19b-5f35-49aa-a53c-34fb024a4331/v1/synthesize?voice=es-ES_EnriqueV3Voice

Fichero SSML.json:

{

"text" : "<prosody rate=\"+15%\" pitch=\"+3st\">Hola Jenaro. <s>Los siguientes quince días vas a estar disfrutando de tu casa. Debido al tamaño que tiene…</s> <break strength=\"medium\"></break> ¿Quieres que te habilitemos un servicio de autobús o medio de transporte para ir de un extremo a otro?.</prosody>“

}

Otros Tags de SSML

- Transformaciones adicionales…

```
<voice-transformation type="Young" strength="80%">
  Could you provide us with new information?
</voice-transformation>

<voice-transformation type="Soft" strength="60%">
  Could you provide us with new information?
</voice-transformation>
```

- Velocidad, fuerza

https://cloud.ibm.com/docs/text-to-speech?topic=text-to-speech-elements

Más ejemplos

<prosody rate="+15%" pitch="+3st">Hola Jenaro. <s>Los siguientes quince días vas a estar disfrutando de tu casa. Debido al tamaño que tiene…</s> <break strength="medium"></break>¿Quieres que te habilitemos un servicio de autobús o medio de transporte para ir de un extremo a otro?</prosody>

https://cloud.ibm.com/docs/text-to-speech?topic=text-to-speech-elements

Opciones adicionales: Entrenamiento de modelos

- Se puede entrenar el servicio (sobre un modelo de lenguaje de terminado) para incluir la lectura de palabras, indicándose, por ejemplo:
  - Si debe ser deletreada letra a letra o números (IBM, SAP, N-323)
  - Incluyendo su descripción en forma de fonemas (Tags de SSML de IPA o propietario de IBM-SPR):

```
<phoneme alphabet="ipa" ph="təmˈɑto"></phoneme>
<phoneme alphabet="ibm" ph="1gAstroEntxrYFXs"></phoneme>
```

Fin