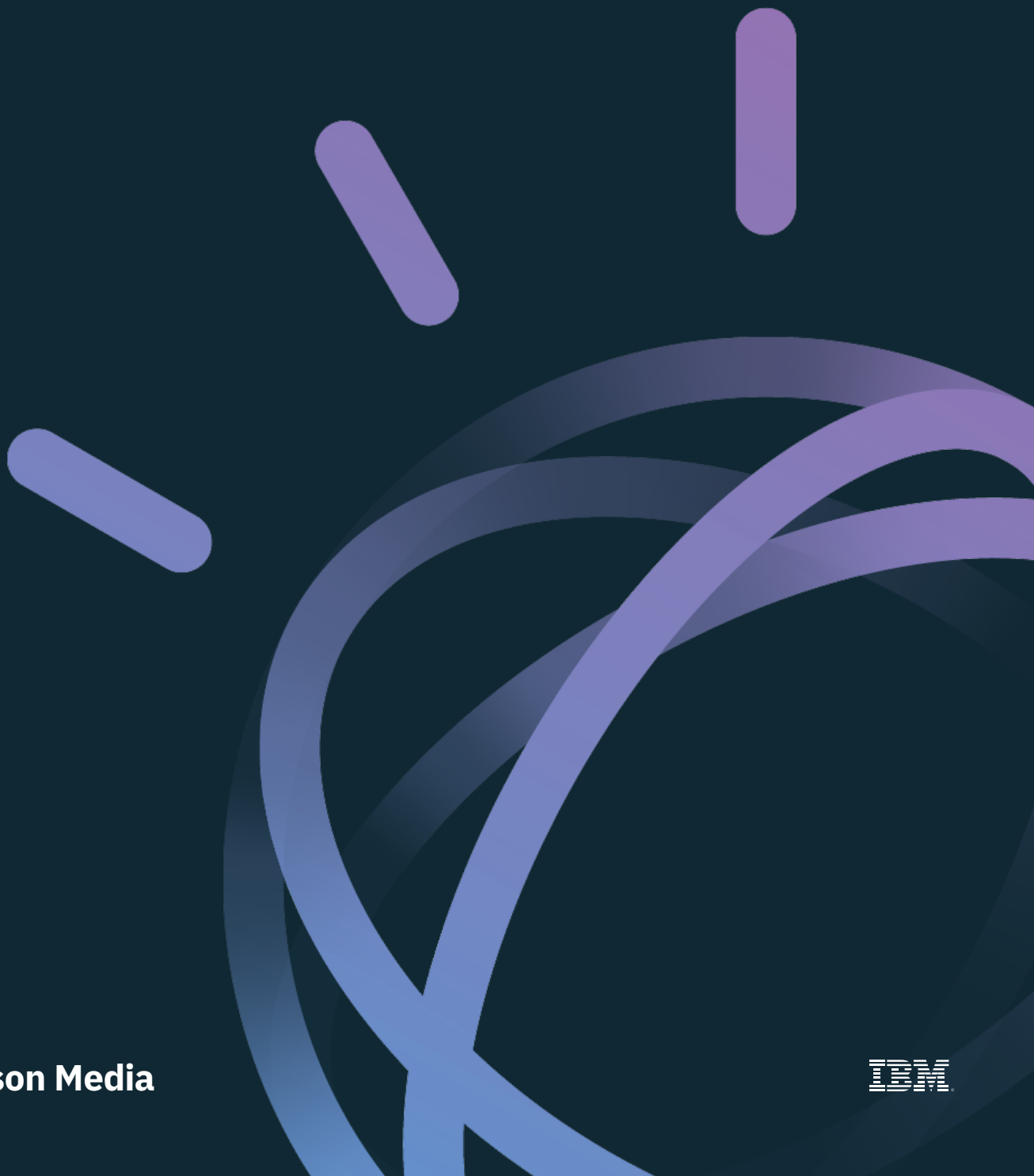
Libro blanco

Watson at Work

La subtitulación se vuelve cognitiva: un nuevo enfoque de un antiguo reto



IBM **Watson Media**

IBM

Los sistemas cognitivos prometen inyectar mayor contexto y mejor precisión a costes competitivos

Asunto:
La subtitulación de vídeos sigue planteando retos. Las nuevas regulaciones pueden incluso intensificar la dificultad.

Contexto:
Incluso con ayuda de sistema automatizados y de especialistas externos, la subtitulación en vivo y en posproducción es un arte relativamente imperfecto que se resiste testarudamente a los avances. Los nuevos requisitos para la subtitulación de cierto contenido de vídeo en línea aumentarán la demanda de subtitulación económica y precisa.

Solución:
Los sistemas cognitivos tienen potencial para aportar nuevas posibilidades a la subtitulación, concretamente en áreas de rápida ingesta y en la aplicación de una comprensión contextual similar a la de los humanos que reduce los errores y mejora la inteligibilidad de los subtítulos.

"Los sistemas cognitivos pueden tener éxito donde las plataformas anteriores de automatización de la subtitulación han demostrado sus carencias".

El primer ejemplo de subtitulacion “cerrada” reconocido ampliamente fue la retransmisión en 1972 del conocido programa de cocina "The French Chef" de la difunta Julia Child en WGBH-TV. Casi desde entonces, han sido muchas las iniciativas presentadas para automatizar o al menos facilitar el meticuloso proceso de traducción de palabras orales y sonidos en subtítulos y descripciones textuales.

A los servicios de subtitulación externos que hacían frente a la demanda de los canales y emisoras les siguió la presentación de software destinado a reconocer sonidos y exposiciones orales. Para demandas más urgentes, la subtitulación en tiempo real apareció en 1982 cuando el Instituto Nacional de Subtitulación inauguró un servicio que prometía la interpretación textual de palabras y sonido en entre cuatro y cinco segundos. Su ingrediente secreto: hordas de reporteros de tribunales con gran agilidad en los dedos y auriculares produciendo maquinalmente texto casi en directo en entre cuatro y cinco segundos.

Todos estos entrantes representaban lo mejor de la vanguardia de la subtitulación. Ninguno era ideal. La subtitulación "en vivo" dio varios pasos en falso ya que los encargados de teclear los textos tenían que publicar inmediatamente palabras, nombres y expresiones no familiares que a veces comprendían y traducían solo con dificultad. Las soluciones de software también han sido imperfectas históricamente y generalmente por los mismos motivos. Las faltas de ortografía, las designaciones "inaudibles" y, especialmente, la dificultad para interpretar el contexto circundante de las palabras orales han puesto a prueba la paciencia de los clientes del sector de la televisión, que están obligados a revisar y corregir los defectos de la producción antes de aplicar los subtítulos a sus vídeos acabados.

Por lo tanto, es fácil simpatizar con un cierto cansancio recurrente entre los profesionales del sector del vídeo, a los que se les han prometido soluciones a prueba de fallos, solo para encontrar lo que siempre ha sido: un proceso manual intensivo y testarudamente imperfecto.

Avance de los sistemas cognitivos

Hoy, sin embargo, existe una nueva posibilidad en la escena de la subtitulacion que tiene el potencial de constituir un auténtico avance. Los sistemas cognitivos combinan la inteligencia ya existente en el tratamiento de los vídeos con algo nuevo: la capacidad para analizar, comprender y "conocer" el contexto del contenido del vídeo de manera muy similar a como lo hacen los humanos. Por lo tanto, la palabra "falta" en un partido de tenis se trata de manera muy diferente a como se trataría en un episodio de una telenovela de emisión diurna. Y la interpretacion y exposición resultantes de las palabras y descripciones anteriores y posteriores son aún más precisas y con una presentación contextual mucho mejor. Los sistemas cognitivos tienen la posibilidad de tener éxito donde las anteriores plataformas de automatización de la subtitulación han demostrado sus carencias porque crean contenido que recoge de forma más fiel la intención, el objetivo y la exposición literal de las palabras y sonidos ligados al contenido del vídeo. Por su capacidad para examinar e interpretar, funcionan de forma casi idéntica a como lo hacen los transcriptores humanos. Excepto que son más rápidos.

"El autoaprendizaje con cada corrección, la precisión del reconocimiento mejora con cada uso".

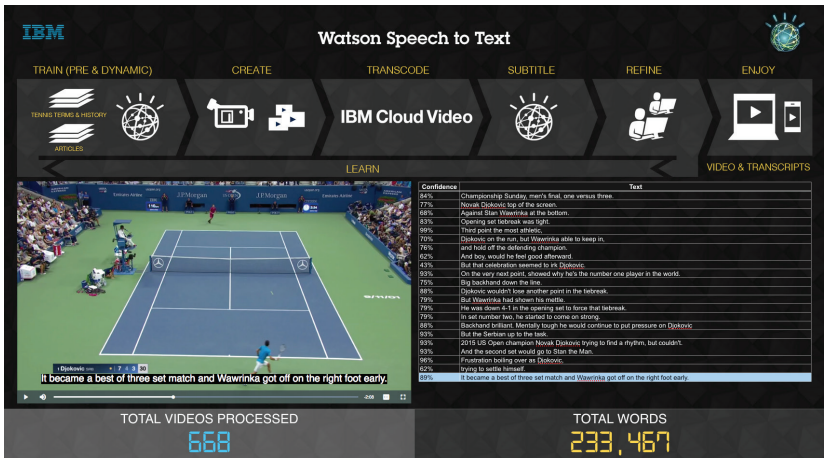
Las tecnologías claves que aporta Watson a efectos de la subtitulación son:

- **La personalización del modelo de lenguaje** crea modelos de lenguaje específicos del sector que mejoran la precisión del reconocimiento
- **El corpora personalizado** amplía el vocabulario con palabras en contexto
- **Las palabras personalizadas** amplían el vocabulario con palabras y su forma fonética
- **Los modelos acústicos personalizados** mejoran la precisión del reconocimiento en vídeos con condiciones de sonido concretas (ruido de fondo, acentos especiales, etc.)

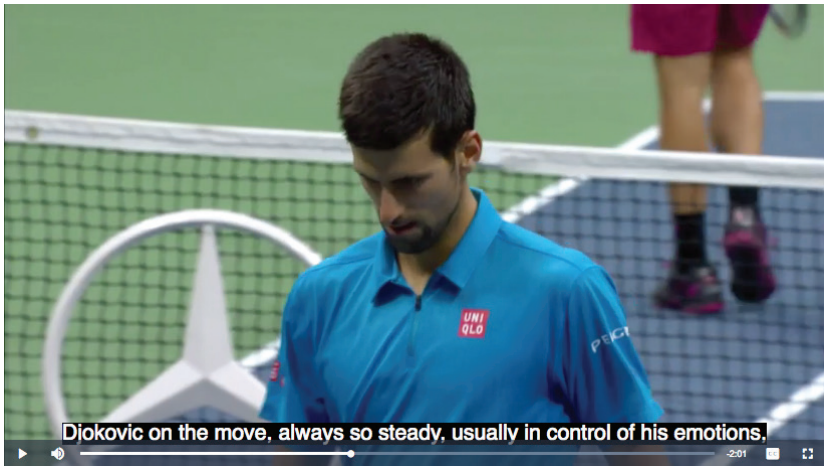
Subtitulación ofrecida por Watson

Watson de IBM utiliza el reconocimiento automático del habla para el tratamiento de los elementos hablados y sonoros de los activos de vídeo. Luego aplica una serie de funciones cognitivas para estimar y actuar en base a los datos interpretados. Además, Watson habilita soluciones de subtitulación personalizada aprovechando características como el corpus, el vocabulario y modelos de audio personalizados para aumentar más aún la precision de los guiones de subtitulación de estreno.

Watson genera automáticamente subtítulos para vídeos utilizando la API Speech to Text de Watson. El editor de subtitulación de la API está diseñado para revisar y corregir los subtítulos generados automáticamente. La interfaz del editor está diseñada para expertos y para no profesionales y está optimizada para lograr una máxima eficiencia. Aprendizaje automático con cada corrección, la precisión del reconocimiento mejora con cada uso. Los nombres y los sustantivos se extraen automáticamente de los subtítulos revisados y se utilizan como contenido de un glosario, garantizando su reconocimiento y su correcta ortografía en usos posteriores. La generación de subtítulos se realiza utilizando un algoritmo de composición inteligente. Con este algoritmo, Watson segmenta automáticamente cada subtítulo por sus puntos de corte naturales, agilizando así su lectura.



La demostración Watson + Abierto de Estados Unidos muestra cómo se generan automáticamente los subtítulos utilizando la API Speech to Text de Watson



Watson subtitula el Abierto de Estados Unidos (demo)

La cuestión del dinero

En última instancia, por supuesto, los sistemas cognitivos tendrán que demostrar su ventaja económica produciendo contenido utilizable a un coste igual o inferior al de la norma que prevalece en el sector. En el rango alto, los gastos en subtitulación de vídeo con contenido en directo se encuentran entre 3 y 5 USD por minuto, incluyendo tanto los guiones generados automáticamente como la edición por parte de seres humanos; en el rango bajo para contenido que no sea no en directo y de vídeo bajo demanda está cerca de 1 y 2 USD por minuto. Estos son los límites que deben mantener o mejorar los sistemas cognitivos para conseguir cualquier tipo de tracción en el mercado.

Las buenas noticias en este sentido es que los sistemas cognitivos participan en un mercado creciente para las herramientas de reconocimiento automático de contenido (o ACR, por sus siglas en inglés) en general, con implicaciones positivas para la economía de escala. La empresa internacional de investigación de mercado MarketsandMarkets estima una tasa de crecimiento anual compuesto de la tecnología ACR del 27,2 % hasta 2021 en el sector de la prensa y el entretenimiento, movida por una confluencia de aplicaciones ligadas al uso de la huella digital de sonido, las marcas de agua y el reconocimiento y el descubrimiento musical. Básicamente, la creciente demanda de formas mejores de encontrar y descubrir contenido de prensa personalizado se prestará a una mayor puesta en común de los costes de las tecnologías ACR en más sectores, lo que puede resultar en la mejora de los costes para los usuarios de la subtitulación (entre otros muchos adoptantes). También contribuye en mayor escala la ampliación del propio mercado de subtitulación de vídeo. Las nuevas reglas estadounidenses para la subtitulación vigentes desde julio de 2017 exigen que el contenido de vídeo en línea que haya sido inicialmente "emitido en TV" aparezca con subtítulos. Este requisito de nuevo puede ampliar el mercado de la subtitulación, aumentando potencialmente las fuentes de inversión y reduciendo los costes.

Watson utiliza el reconocimiento automático de la voz para ingestar elementos hablados y sonoros de los activos de vídeo.

Precisión, precisión, precisión

Por supuesto, el otro gran punto a tener en cuenta para los actores del sector del vídeo es el cumplimiento de las leyes federales. En EE. UU., la FCC exige que la subtitulación sea:

- **Precisa:** Los subtítulos deben corresponderse con lo hablado en el diálogo y transmitir los ruidos de fondo y demás sonido en la mayor medida posible.
- **Síncrona:** Los subtítulos deben coincidir con su correspondiente diálogo hablado y los sonidos en la mayor medida posible y deben aparecer en pantalla a una velocidad que permita que los espectadores los lean.
- **Completa:** Los subtítulos deben aparecer desde el principio al final del programa en la mayor medida posible.
- **Correctamente situada:** Los subtítulos no deben bloquear otro contenido visual importante en la pantalla, solaparse entre sí ni salirse del extremo de la pantalla de vídeo.

Los sistemas cognitivos pueden cumplir la mayor parte de estos requisitos básicos gracias a su habilidad para acoplar el reconocimiento del audio con una mayor comprensión contextual, de modo que las transcripciones presentadas a los editores para su revisión final sean más precisas y estén en mejores condiciones para salir al mercado de lo que permitían las tecnologías anteriores.

Escenarios de la implementación

Los participantes del sector del vídeo que quieren comprender la contribución que pueden realizar los sistemas cognitivos a la subtitulación en el futuro pueden querer experimentar con implementaciones en un estado inicial a modo de prueba. Las circunstancias que pueden dar lugar a las pruebas iniciales pueden incluir cualquiera de estas consideraciones:

- Cuando el coste total es similar o inferior al de las soluciones disponibles
- Cuando se prefiere una solución interna
- Cuando el tiempo de entrega es un factor clave
- Para contenido no sujeto a regulaciones

La función de los sistemas cognitivos en las prácticas de subtitulación televisiva está en sus primeras fases. Pero hay un interés evidente en adoptar una solución que prometa trascender las limitaciones y los límites de los enfoques heredados. La potente combinación del reconocimiento automático del contenido con capacidades cognitivas/de aprendizaje aportará nuevas capacidades a una práctica ya antigua en el mundo de la televisión. Al final, el conocimiento puede ser exactamente lo que quería la subtitulación ya desde que Julia Child mostró al mundo cómo cocinar como un master chef francés.

IBM Cloud Video
550 Kearny Street, Suite 600
San Francisco, CA 94108

Fabricado en Estados Unidos de América
Diciembre de 2017

Este documento está vigente a partir de la fecha inicial de publicación y puede ser modificado por IBM en cualquier momento. No todas las ofertas se encuentran disponibles en todos los países en los cuales IBM opera.

La información de este documento se proporciona "tal cual", sin ninguna garantía, explícita o implícita, incluidas, sin limitaciones, las garantías de comercialización e idoneidad para una finalidad concreta y cualquier garantía o condición de no infracción.

Los productos de IBM están garantizados de acuerdo con los términos y condiciones de los acuerdos bajo los que se proporcionan.

Descripción de Prácticas Recomendadas de Seguridad: la seguridad de los sistemas de TI implica la protección de los sistemas y la información a través de la prevención, la detección y la respuesta frente al acceso indebido desde el interior y el exterior de la empresa. El acceso indebido puede comportar información alterada, destruida o apropiada indebidamente, o puede suponer el daño o mal uso de los sistemas del Cliente para atacar a otros usuarios. Sin un enfoque global de seguridad, ningún sistema o producto de TI puede hacerse completamente seguro y ningún producto o medida de seguridad puede ser totalmente eficaz en la prevención del acceso indebido. Los sistemas y productos de IBM están diseñados para formar parte de un enfoque de seguridad integral, que necesariamente implicará procedimientos adicionales de funcionamiento y podrá requerir que otros sistemas, productos o servicios sean más eficaces.

IBM no garantiza que los sistemas y productos sean inmunes ante conductas malintencionadas o ilegales de alguna de las partes.