

Inteligencia artificial y datos masivos en archivos digitales sonoros y audiovisuales

Perla Olivia Rodríguez Reséndiz
Coordinadora



Q335 Inteligencia artificial y datos masivos en archivos digitales
I57 sonoros y audiovisuales / Coordinadora Perla Olivia Rodríguez
Reséndiz. - México: UNAM. Instituto de Investigaciones
Bibliotecológicas y de la Información, 2020.

xviii, 182 p. - (Tecnologías de la información)

ISBN:

Investigación realizada gracias al programa

DGAPA - PAPIIT IT400118.

1. Inteligencia artificial - Procesamiento de datos. 2. Internet
de las cosas. 3. Archivos sonoros. 4. Big data. I. Rodríguez
Reséndiz, Perla Olivia, coordinadora. II. ser.

Diseño de portada: Oscar Fernando Arcos Casañas

Imágenes:

Envato Elements

(<https://elements.envato.com/es-419/>)

Primera edición, 2020

D.R. © UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Ciudad Universitaria, 04510, México D.F.

Impreso y hecho en México

ISBN: En trámite

Publicación dictaminada

Reflexiones en torno a la inteligencia artificial: el caso de los archivos de RTVE

VIRGINIA BAZÁN-GIL

Responsable de proyectos Fondo Documental RTVE

INTRODUCCIÓN

El objetivo de mi comunicación no es otro que el de dar a conocer la travesía de los archivos de RTVE hacia la generación automática de metadatos. Es un viaje sin duda repleto de quimeras, espejismos y peligros, al final del cual, la tierra prometida parece esperarnos. Porque sin duda, una vez alcanzada la digitalización de los archivos, la inteligencia artificial es lo primero en lo que pensamos cuando alguien nos dice “pide un deseo”.

En primer lugar, presentaré el Fondo Documental de RTVE, a continuación, y de forma muy breve, expondré las tecnologías implicadas en la generación automática de metadatos, después señalaré su utilidad en la cadena de producción, difusión y archivo de los contenidos, identificaré a los proveedores tecnológicos y presentaré algunos casos de uso de la tecnología en las que hemos trabajado. Finalmente, les plantearé algunas conclusiones, poco definitivas, sobre el largo viaje del archivo de una televisión pública hacia la generación automática de metadatos.

EL FONDO DOCUMENTAL DE RADIO TELEVISIÓN ESPAÑOLA

El archivo de Radio Televisión Española es un archivo dividido en dos: el de radio y el de televisión. El archivo de Radio Nacional de España

(RNE) fue establecido en 1957 y conserva los programas emitidos por nuestras seis estaciones radiofónicas. El primer archivo sonoro de producción propia que conservamos data de 1947, se trata de los discursos oficiales de Franco y Eva Perón durante el viaje oficial de la primera dama argentina a nuestro país. El archivo de televisión ha cumplido ya 63 años. La grabación más antigua de producción propia se remonta al año 1959 y corresponde a un rodaje en las instalaciones de Miramar en Barcelona. Este archivo conserva, además, otras colecciones como los noticiarios y documentales de NO-DO producidos entre 1943 y 1981.

La principal misión del fondo documental de RTVE es la preservación del archivo, con el fin de garantizar su disponibilidad para la producción, reutilización y la comercialización de los contenidos. Además, como servicio público estamos comprometidos con la accesibilidad de nuestras colecciones, ya que preservamos contenidos de un indudable valor histórico. Pero, ¿cómo preservamos ese patrimonio? El archivo de RNE se digitalizó de forma temprana en 2002. La digitalización del archivo de TVE se realizó entre 2008 y 2014, con la colaboración de Telefónica. Este proyecto fue sin duda el mayor de su tipo en la digitalización de un archivo de televisión, implementado en España (Lopez-De-Quintana-Saenz 2014). En la actualidad esta tarea se hace con medios propios y queda únicamente por digitalizar el material de los centros territoriales y nuestro archivo de cine, una verdadera joya que requiere de procesos artesanales de recuperación de forma previa a la digitalización.

Como parte de nuestro plan de preservación digital hemos llevado a cabo la migración de nuestras cintas de datos de LTO4 y LTO5 a LTO7 y se ha multiplicado por cuatro la capacidad de nuestro almacenamiento digital. Asimismo, hemos logrado la unificación de las librerías de programas e informativos, de forma que podremos gestionar de manera más eficiente nuestros recursos.

La gestión de las colecciones (texto, audio, video y fotografías) se realiza a través del gestor documental ARCA, un desarrollo propio implantado en 2011, que permitió la unificación de las distintas bases documentales y la adopción de un modelo de metadatos común, tanto para el archivo de radio como para el de televisión y las distintas

unidades implicadas. ARCA está integrado con otras aplicaciones de negocios, como emisiones, comercial, subtítulo, unidades de producción y RNE, así como con el CMS de medios interactivos, lo que permite a nuestros compañeros de la web publicar contenidos del archivo.

Como servicio público estamos obligados a abrir nuestras colecciones a los usuarios pero al mismo tiempo tenemos que cubrir nuestras necesidades de producción, lo que implica un equilibrio inestable de recursos, necesidades y demandas. Estamos realizando un gran esfuerzo para publicar el archivo de RTVE en la web (RTVE 2019) y responder a las peticiones de los investigadores. En este sentido, estamos explorando nuevas vías, la Biblioteca Nacional Española (BNE), con quien además estamos trabajando en una política de preservación de nuestros contenidos en la web y de quienes esperamos obtener, en el futuro, la calificación de repositorio seguro. Tenemos convenios de colaboración con distintas instituciones patrimoniales, entre otras, con la Filmoteca Española. Somos miembros activos de la Federación Internacional de Archivos de Televisión (FIAT/IFTA 2019) y colaboramos con otras iniciativas europeas como EUScreen (EUScreen 2019).

TECNOLOGÍAS DEL HABLA, PROCESAMIENTO DEL LENGUAJE NATURAL Y VISIÓN ARTIFICIAL

Los procesos de generación automática de metadatos en los archivos se fundamentan en tres tecnologías complementarias, visión artificial, tecnologías del habla y procesamiento del lenguaje natural (Lleida 2018):

- La visión artificial se ocupa del reconocimiento de imágenes, la agrupación y segmentación de escenas y el seguimiento de objetos y personas.
- Las tecnologías del habla, permiten tanto la conversión voz a texto como un tratamiento de los hablantes y sus emociones, facilita su identificación y la segmentación del discurso.

- El procesamiento del lenguaje natural permite el reconocimiento de entidades y temas tratados en un discurso y facilita la creación de resúmenes y descripciones.

En otras palabras, la visión artificial proporciona la descripción de la capa de imagen, bien en forma de etiquetas, bien en forma de descripciones, en lenguaje natural (Bazán-Gil y Guerrero Gómez-Olmedo 2018). Las tecnologías del habla permiten convertir la capa de audio en un texto y puede aplicarse tanto en la fase de edición para la transcripción automática de entrevistas, por ejemplo, como en la emisión y difusión de contenidos través de la web en forma de subtitulado; y finalmente, el procesamiento del lenguaje natural contribuye a enriquecer con metadatos el material en la publicación web y en el archivo, al generar puntos de acceso en forma de entidades para la recuperación o al facilitar la clasificación automática de los contenidos con el uso de ontologías. (Bazán-Gil et al. 2019)

PERO, ¿QUIÉN NOS PROPORCIONA LA TECNOLOGÍA?

Podemos identificar tres grupos de proveedores de inteligencia artificial:

- Grandes compañías internacionales Microsoft Azure (Microsoft 2019), IBM Watson (IBM 2019) y Google AI (Google 2019), mismas que gestionan una gran cantidad de datos procedentes de *social media*. Tecnología propietaria en constante proceso de mejora.
- Compañías especializadas en el sector *media*, que integran soluciones de terceros o tienen sus propios desarrollos basados en tecnologías *open source*. Pueden ser MAM Vendors como: AVID (AVID 2019), Dalet (DALET 2019), VSN (VSN 2019), o Tedial (TEDIAL 2019), o bien empresas dedicadas como Etiqu-media (ETIQMEDIA 2019) y Limecraft (Limecraft 2019), que se ocupan únicamente del procesamiento de los contenidos en flujos secundarios.

- Grupos de investigación nacionales e internacionales. Esta es una fórmula de colaboración que en el caso de RTVE ha dado buenos resultados, ya que si algo demanda la IA es el procesamiento de grandes cantidades de datos, algo en lo que somos ricos los archivos. Identificar grupos de investigación en tecnologías del habla, procesamiento del lenguaje natural e inteligencia artificial y crear cátedras o solicitar proyectos de investigación conjuntos, puede ser una fórmula de éxito.

PERO, ¿ES LA GENERACIÓN DE METADATOS PATRIMONIO EXCLUSIVO DEL ARCHIVO?

La generación de metadatos no ha sido nunca patrimonio exclusivo del archivo, tradicionalmente otras áreas como las de producción, emisión o comercialización han generado metadatos para los mismos contenidos audiovisuales en función de sus propias necesidades. Y en el caso de RTVE, de forma reciente, estamos experimentando la gran demanda de datos procedente del área de medios interactivos. Los sitios web de las radiodifusoras demandan metadatos para identificar, clasificar, enriquecer, hacer más accesibles y buscables los contenidos y esta demanda se produce a un ritmo mayor del que los propios archivos podemos asumir. Más datos y más rápido. Y a ninguno de nosotros como profesionales se nos escapa que a menudo estas áreas cuentan con la capacidad de acción y los recursos necesarios para emprender nuevos proyectos tecnológicos que aun siendo necesarios son impensables para el archivo. Y esto, que puede considerarse como una amenaza, no es sino una oportunidad para encontrar en estas áreas aliados estratégicos en la adopción de las nuevas tecnologías. Porque como repetía el slogan “*Solo no puedes, con amigos sí*”, del mítico programa de TVE *La bola de cristal*, de los años 80.

Aún más, comprender el ciclo de vida de los datos en el proceso de producción, difusión y archivado, no solo nos permite identificar estas alianzas, sino que puede contribuir a la eficiencia y sostenibilidad de nuestros datos como organización (Gomez Zotano y Bazán-Gil 2018) Llamémoslo si quieren la ecología del dato ¿por qué producir constantemente nuevos metadatos cuando podemos usar los que otros ya han generado?

Esta es una de las lecciones más valiosas que hemos aprendido, prueba de concepto tras prueba de concepto, en los últimos tres años.

CASOS DE USO DE LA TECNOLOGÍA EN RTVE

Una vez identificadas las tecnologías y su utilidad, los *partners* tecnológicos y los aliados estratégicos, repasaremos algunas de las experiencias en el uso de la AI en Radio Televisión Española. En concreto, las más recientes que han tenido como protagonistas las tecnologías del habla: Iberspeech 2018 (Lleida et al. 2019) y el proyecto “Crónicas”.

Dentro de las tecnologías del habla, las más relevantes para nosotros son: transcripción de voz a texto (Speech to Text) y segmentación y agrupación por hablantes, también conocida como diarización.

Los sistemas de voz a texto nos permiten conocer qué se ha dicho y son fundamentales para lograr:

- Garantizar la accesibilidad de los contenidos mediante la generación de subtítulos en tiempo real y la posibilidad de traducción automática a otros idiomas.
- Mejorar los tiempos de producción a través de la transcripción automática de entrevistas, ruedas de prensa, etc.
- Automatizar parcialmente los procesos de análisis, ya que en las transcripciones se basa la clasificación y etiquetado automático contenidos.

La diarización nos ayudará a identificar quién habla, cuándo habla y qué aspecto tiene si aplicamos, además, técnicas de reconocimiento facial:

- Indexación de hablantes y discursos.
- Estructuración de contenidos.
- Reconocimiento de hablantes como apoyo al Speech to Text.

IBERSPEECH

En este contexto, se impulsó la investigación en tecnologías del habla en español, en mayo de 2018, con la Catedra RTVE Universidad de

Zaragoza (Cátedra RTVE Universidad de Zaragoza 2017), instituida un año antes con el objetivo de investigar en la creación automática de metadatos. De esta manera, RTVE Iberspeech Challenge 2018 (RTVE 2018), lanzó un reto tecnológico que puso a disposición de la comunidad científica más de 500 horas de contenidos emitidos, junto con sus correspondientes subtítulos.

El conjunto de datos liberado incluye cerca de una veintena de programas de distintos géneros y temática, producidos y emitidos por RTVE, entre 2015 y 2018. Estos programas presentan distintas dificultades desde el punto de vista de las tecnologías del habla como son: la diversidad de acentos del español, la superposición de diálogos, el habla espontánea, la variabilidad acústica, el ruido de fondo o el vocabulario específico, factores todos ellos que inciden en el rendimiento de los sistemas de transcripción.

Sobre este conjunto de programas, los 22 grupos participantes debían detectar automáticamente, etiquetar y transcribir segmentos del habla. Para ello se definieron tres tareas distintas: voz a texto, diarización y diarización multimodal.

RETO 1: TRANSCRIPCIÓN VOZ A TEXTO

El objetivo de este reto era transcribir de forma automática 39 horas de emisión. La métrica aplicada, para medir el rendimiento de los sistemas en la WER o tasa de error por palabras, que tiene en cuenta el número de inserciones, sustituciones y borrados en relación con el número total de palabras (Lleida et al. 2019). Los resultados de este reto ponen de manifiesto que la calidad de la descripción depende del tipo de programa:

- El sistema mejor evaluado obtiene una tasa de error del 16.45% y el peor evaluado un 35.80%
- Programas con entrevistas o conversaciones espontáneas y con audio en exteriores como “Dicho y Hecho”, “La Mañana” o “Arranca en verde” son los programas que suponen el mayor reto con tasas de error muy por encima del 20% para la mayoría de los sistemas.

- Muchos programas mantiene unas tasas de error por debajo del 20% e incluso para un programa como “Latinoamérica en 24H”, con distintos acentos del español, la tasa de error se sitúa por debajo del 10%.

RETO 2: DIARIZACIÓN

El objetivo de este reto era segmentar y agrupar por hablantes 22 horas de emisión. La tasa de error que se aplicó para medir el rendimiento de los sistemas es el DER o tasa de error en la diarización que tiene en cuenta el número de segmentos que no se atribuyen correctamente a un hablante (Lleida et al. 2019). Los resultados de este reto demostraron:

- La variabilidad en las prestaciones en función del tipo de contenido (tasas de error, entre el 17.27% y el 39.09%) .
- La asignación incorrecta de hablantes como error más común.

RETO 3: DIARIZACIÓN MULTIMODAL

El objetivo de este reto era segmentar y agrupar por hablantes y cara cuatro horas de emisión. La tasa de error aplicada es la correspondiente en la diarización por hablantes y caras. Los resultados demuestran la variabilidad en función del tipo de programa y un mejor rendimiento de la diarización de hablantes frente a la de caras (Lleida et al. 2019).

Estos resultados deben abordarse además desde la perspectiva del usuario, es decir desde la utilidad de las tecnologías para su integración en los flujos de trabajo de una organización como RTVE.

En los últimos cinco años los reconocedores de voz han mejorado de forma espectacular. Aun así, todavía existe una gran variabilidad en la tasa de error, que limita su utilidad en todos los escenarios posibles: desde el material original hasta los programas emitidos.

EL PROYECTO “CRÓNICAS”

Este proyecto, también en el contexto de las tecnologías del habla pretendía dar respuesta a un largo anhelo de los periodistas, la trans-

cripción automática de las entrevistas que realizan para sus reportajes. El objetivo fundamental de esta prueba piloto era determinar la utilidad de los reconocedores automáticos en la producción de nuevos contenidos. Para ello, quisimos probar distintas herramientas, su posible integración con nuestra herramienta de producción y determinar que tasa de error era asumible para nuestros usuarios. Este proyecto se realizó en dos fases. En la primera fase se definieron los requisitos del sistema, en colaboración con el equipo del programa, y se identificaron como fundamentales los siguientes:

- Sistema intuitivo y fácil de usar.
- Múltiples usuarios conectados al mismo tiempo
- Debe conservar los códigos de tiempo originales de la grabación
- La segmentación de la entrevista debe realizarse por hablantes y diferenciar entre preguntas y respuestas.
- Las transcripciones deben incluir capitalización y signos de puntuación.
- La edición y corrección de la transcripción generada debe ser sencilla.
- Tasas de error no superiores al 15%.

En esta primera fase, también, se definió el alcance de la muestra para realizar las pruebas. Se seleccionaron 25 rodajes originales con base en factores como la variedad de modelos del lenguaje, la calidad de audio, la presencia de habla espontánea, la existencia de diálogos superpuestos y de diferentes acentos del español.

Durante la segunda fase de este proyecto se llevaron a cabo pruebas con dos soluciones tecnológicas: MAM VSN y Limecraft.

El MAM de VSN integra varias soluciones de IA de Microsoft Azure y Google Cloud, para esta prueba se usaron ambas, con el objetivo de establecer una comparación entre el rendimiento de cada una. Los ficheros se subieron al MAM de VSN, mediante un sistema de *drag and drop* que permiten arrastrar directamente los ficheros desde el equipo del redactor a la solución de VSN. Los tiempos de subida y procesamiento de los ficheros, uno de los puntos críticos del proyecto, fueron satisfactorios -menos de la mitad de duración del fichero,

en el procesamiento, no así las tasas de error devueltas por ambos sistemas. Tanto Google como Microsoft, en todos los rodajes, devolvieron tasas de error por encima del 50%, además, destaca el hecho de que los ficheros con audio en alta calidad procesados por Google presentaron tasas de error más altas que los ficheros en baja calidad.

Tanto Azure como Google generan segmentos en torno al minuto de duración, es decir, no segmentaron por hablantes o al menos por fragmentos del discurso con significado semántico. Google no devuelve signos de puntuación ni capitalización y Azure sí lo hace, pero tiende a finalizar casi todas las frases con un signo de interrogación.

Una segunda iteración se llevó a cabo con Limecraft (Limecraft 2019), una solución cloud basada en Speechmatics (Speechmatics 2019), un reconocedor del habla con muy buenos resultados en distintos retos internacionales y que integran distintos canales de televisión internacionales. Es por tanto una herramienta específica de transcripción de voz a texto. Limecraft presentaba una interfaz sencilla de uso, proporcionaba segmentación por hablantes, capitalización y signos de puntuación y la posibilidad de editar los contenidos directamente en la interfaz de consulta. Las tasas de error para para los distintos rodajes se situaron en torno del 20% al 30%, considerablemente mejores que las obtenidas por Google y Azure y, a diferencia de estos sistemas, Limecraft devuelve segmentación por hablantes, aunque como hemos visto en los resultados del reto de diarización en Iberspeech 2018, este no es un problema totalmente resuelto.

Una vez finalizadas las pruebas y, a pesar de que la percepción de la redacción respecto a la utilidad de la tecnología evolucionó positivamente a lo largo del proyecto, en ningún momento llegó a considerarse su integración de esta tecnología en los flujos de trabajo habituales de la redacción del programa. En opinión de los periodistas que participaron, el esfuerzo en la corrección de las transcripciones distraía su atención de la rememoración de la entrevista y dificultaba el proceso de edición. Nuestros colegas tampoco valoraron positivamente la integración de estas soluciones con herramientas de edición.

Desde el punto de vista del archivo, la integración de estas soluciones en la producción de contenidos es muy relevante, ya que los rodajes originales y, de forma más concreta, las entrevistas, no suelen

llegar al archivo con los metadatos suficientes para su adecuado tratamiento. La transcripción exacta de lo que se ha dicho, cobra especial relevancia en el contexto de la verificación de noticias y sigue siendo relevante para la recuperación en tanto que multiplica los puntos de acceso a los contenidos en la búsqueda por texto libre.

CONCLUSIONES

En los últimos 5 años los sistemas de reconocimiento del habla han mejorado considerablemente sus tasas de error. De acuerdo con las pruebas que hemos realizado en los dos últimos años, las tasas de error pueden situarse entre el 5% y el 50% y esto, demuestra que no hay soluciones universales. Existe una correlación entre tipo de programa y resultado de la transcripción. En este sentido, y desde el punto de vista del archivo, es necesario comprender el funcionamiento de la tecnología y bajo qué premisas tiene un mejor rendimiento. Este conocimiento debe ponerse además en relación con las necesidades reales de trabajo y con las expectativas de las distintas áreas en cuanto a la calidad de los resultados. La inteligencia artificial es enemiga de la calidad humana. No debemos olvidar tampoco que el grado de tolerancia respecto a los errores varía en función del escenario para el que se generan los metadatos: tasas de error por encima del 15% pueden ser tolerables para el archivo cuando se carece de otros datos para la recuperación, pero no son admisibles para la emisión de los contenidos.

Para los archivos audiovisuales la incorporación de IA a sus procesos diarios es más que una prioridad, pero la pregunta es ¿por dónde empezar? Para iniciar este viaje es necesario concienciarse sobre la importancia, no solo de generar metadatos en función de las necesidades de las distintas áreas, sino de establecer los mecanismos adecuados para garantizar la reutilización de estos datos en la cadena de producción - emisión - archivado. Si comprendemos el ciclo de vida de los datos en esta cadena seremos capaces de identificar a nuestros aliados estratégicos para este viaje.

Tras las experiencias y el conocimiento adquirido hasta el momento, podemos identificar algunas acciones como fundamentales antes

de embarcarse en el viaje hacia la implantación de la inteligencia artificial en un archivo de televisión:

- Analizar las necesidades como archivo y comprender las de otras áreas. ¿Dónde es prioritaria la generación de metadatos? Es necesario poner el foco en función de nuestros recursos.
- Identificar proveedores tecnológicos y grupos de investigación relevantes en las tecnologías que se quieran poner a prueba.
- Realizar una prospección de la tecnología, de sus capacidades y especialmente de sus limitaciones.
- Crear sinergias con áreas que puedan impulsar y hacer realidad la adopción de la tecnología: “Solo no puedes, con amigos sí”.
- Examinar el modelo de metadatos actual ¿es suficiente para asumir la avalancha de nuevos datos procedentes de la inteligencia artificial?
- Ajustar las expectativas de los usuarios al rendimiento real de la tecnología.
- Evitar el uso indiscriminado de la tecnología como la única forma posible de malgastar recursos económicos y humanos, sin la garantía de obtener resultados adecuados.

Por último, la adopción de soluciones de inteligencia artificial en distintas áreas de una cadena de radio y televisión es una magnífica oportunidad para hacer visible el archivo. El poder de los datos es propio de nosotros los y las archiveras. No duden en convertirse en el oráculo de otras áreas dentro de la organización.

REFERENCIAS

- AVID. (2019). “Avid - Tecnología y Herramientas Para Capacitar a Los Creadores de Contenido.” <https://www.avid.com/#Video>.

Bazán-Gil, Virginia, and Ricardo Guerrero Gómez-Olmedo. (2018). "Descripción Automática de Archivos Audiovisuales: NeuralTalk, Un Modelo de Video2text Aplicado Al Archivo de RTVE Cita Recomendada." *BiD: Textos Universitaris de Biblioteconomia i Documentació*, no. 41. <https://doi.org/10.1344/BiD2018.41.7>.

Bazán-Gil, Virginia, Eduardo Lleida, Carmen Pérez, Manuel Gómez, and Alberto De Prada. (2019). "Tecnologías Del Habla: Nuevas Oportunidades Para Los Archivos de Televisión." In *IV Congreso ISKO España-Portugal*. Barcelona: Universitat de Barcelona.

Cátedra RTVE Universidad de Zaragoza. (2017). "Cátedra RTVE de La Universidad de Zaragoza." <http://catedrartve.unizar.es/>.

DALET. (2019). "Orchestrated Media Solutions | Dalet." <https://www.dalet.com/>.

ETIQMEDIA. (2019). "Catalogación Supervisada de Contenido Audiovisual. ETIQMEDIA. Inicio. ETIQMEDIA." <http://www.etiqmedia.com/>.

EUscreen. (2019). "EUscreen." <http://www.euscreen.eu/>.

FIAT/IFTA. (2019). "FIAT/IFTA." <http://fiatifta.org/>.

Gomez Zotano, Manuel, and Virginia Bazán-Gil. (2018). "Interoperabilidad y Recomendación de Contenidos: Explotando El Archivo En La Web." II Jornadas Fondo Documental RTVE: Los Archivos de Televisión Después de La Digitalización. <http://www.rtve.es/rtve/20180202/jornadas-archivos-televison/1672420.shtml>.

Google. (2019). "Productos de Inteligencia Artificial y Aprendizaje Automático." <https://cloud.google.com/products/ai/?hl=es>.

- IBM. (2019). "IBM Watson Products and Solutions | IBM." <https://www.ibm.com/watson/products-services>.
- Limecraft. (2019). "Limecraft - Your Video Production Workflow Set up in a Few Easy Steps : Limecraft". <https://www.limecraft.com/>.
- Lleida, Eduardo. (2018). "Tecnologías Para AI Análisis y Metadato de Contenidos Audiovisuales." Los Archivos de Televisión Después de La Digitalización: Gestión y Acceso Abierto. http://www.rtve.es/contenidos/documentos/instituto/4_Jornada_Archivos_tv.pdf.
- Lleida, Eduardo, Alfonso Ortega, Antonio Miguel, Virginia Bazán, Carmen Pérez, Manuel Gómez, and Alberto De Prada. (2019). "Albayzin 2018 Evaluation : The IberSpeech-RTVE Challenge on Speech Technologies for Spanish Broadcast Media." *Applied Sciences*, 1-23. <https://www.mdpi.com/journal/applsci>.
- Lopez-De-Quintana-Saenz, Eugenio. (2014). "Rasgos y Trayectorias de La Documentación Audiovisual: Logros, Retos y Quimeras." *Profesional de La Informacion*. <https://doi.org/10.3145/epi.2014.ene.01>.
- Microsoft. 2019. "Video Indexer: Información de Vídeos de IA | Microsoft Azure." 2019. <https://azure.microsoft.com/es-es/services/media-services/video-indexer/>.
- RTVE. (2018). "La Cátedra RTVE En La Universidad de Zaragoza Presenta Su Primer Reto Tecnológico a La Comunidad Científica - RTVE.Es." <http://www.rtve.es/rtve/20180521/catedra-rtve-universidad-zaragoza-presenta-su-primer-reto-tecnologico-comunidad-cientifica/1737360.shtml>.
- _____. (2019). "Archivo RTVE." Archivo RTVE. <http://www.rtve.es/television/archivo/>.

Speechmatics. (2019). "Speech to Text Recognition Powered by Machine Learning | Speechmatics." <https://www.speechmatics.com/>.

TEDIAL. (2019). "Tedral Media IT | Media Asset Management (MAM) Solutions." <https://www.tedral.com/>.

VSN. (2019). "VSN | Innovation and Media Solutions." <https://www.vsn-tv.com/es/>.

Inteligencia artificial y datos masivos en archivos digitales sonoros y audiovisuales.

Instituto de Investigaciones Bibliotecológicas y de la Información/UNAM. La edición consta de 100 ejemplares. Coordinación editorial, Israel Chávez Reséndiz; revisión especializada, Angélica Valenzuela y Valeria Guzmán González; revisión de pruebas, Valeria Guzmán González; formación editorial, Oscar Fernando Arcos Casañas. Fue impreso en papel cultural de 90 gr. en los talleres de Grupo Fogra. Año de Juárez 223. Col. Granjas San Antonio. Alcaldía Iztapalapa. Ciudad de México. Se terminó de imprimir en 2020.