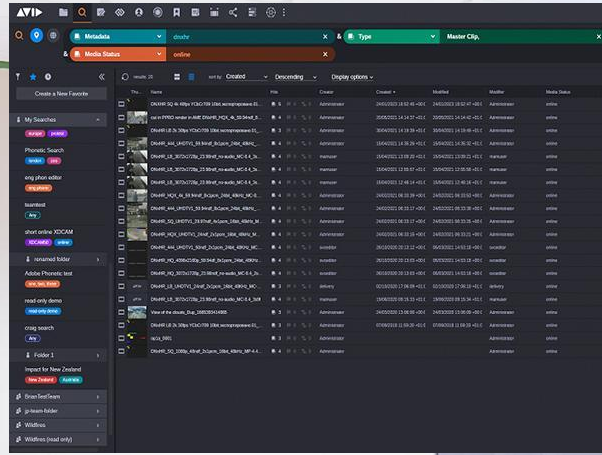


***The RTVE journey towards
applications of AI in the archive
– collaboration with
researchers.***

***Virginia Bazán-Gil
RTVE Archive***



rtve Comercial



rtve play



1,200,000 hours of video

1,300,000 hours of audio



rne audio

In short

- The RTVE Archive strategy for Artificial Intelligence is focused on process improvement, specifically enhancing content accessibility through automatic generation of metadata.
- This strategy has been implemented since 2021 through automatic metadata generation projects awarded to external providers via public tenders.
- Implementation has been useful to **learn how AI fits** into our **workflows**, into the **archive culture**, and into the technological structure of the organisation.
- A transparent **communication policy** has been adopted, along with a continuous training plan for the Archive team.



Sincronizar vídeo con texto Pista 2

Datos del vídeo
Referencia: I01750792
Título: TD2 CR 20241202
Versión del software de análisis: v090

Resumen
El Tribunal Supremo citó como imputado al exministro José Luis Ábalos para el 12 de diciembre. También ha llamado a declarar a Víctor Valdama y Coldo García en el caso Coldo. En Francia, el primer ministro Barnier enfrenta una moción de censura debido a la falta de apoyos para los presupuestos. El Tribunal ha permitido que la bandera LGBTI ondee en edificios públicos, argumentando que no es un símbolo partidista. En el ámbito sanitario, dos lesionados medulares han recuperado la capacidad de andar tras estimular el hipotálamo. Glovo decidirá dejar el modelo de falsos autónomos y contratará a sus repartidores antes de un juicio a su fundador por delitos laborales. En el PSOE están en proceso de renovación de liderazgos tras la reciente unión de su ejecutivo. En Estados Unidos, el presidente Biden

Una hora aproximadamente, Ana, muy buenas noches, de reunión entre Hans y Flick y la plantilla a la que también se ha unido el presidente Laporta después de perder contra las palmas para cerrar un mes de noviembre para olvidar y que ha apretado la liga y de qué manera. Hace tres jornadas el Barça solo había perdido una vez, acababa de golear en el Bernabéu y le sacaba a Madrid y Atlético nueve y diez puntos respectivamente.

00:32:38.073 - 00:32:50.093 Hablante: #s37

Pero este lunes la clasificación amanece así, los culés apenas han sumado un punto en tres partidos, el Madrid se ha puesto a uno y con el partido de Valencia pendiente y el Atlético tras su goleada está solo a dos.

00:32:51.273 - 00:33:24.093 Hablante: #s17

Lo bueno es que noviembre se ha acabado, dice sonriendo Flick, su Barça muestra la versión más terrenal. Somos capaces de crear ocasiones pero no de definir palo tras palo, 14 llevan, los que más de las cinco grandes ligas. El equipo más joven de la liga, contextualiza, hemos flaqueado un poco en la presión como bloque. La solución, luchar, no se trata solo de jugar al fútbol y esto es muy importante. Este martes visita al Mallorca, sexto y tras dos triunfos seguidos.

00:33:26.313 - 00:33:59.373 Hablante: #s17

Es el momento de decir, aquí estamos nosotros. Tres jornadas sin ganar incluidas, dos derrotas acercan al renacido Real Madrid, que ya le mira a los ojos. Tres

Palabras clave

Entidades (NER)

Fechas

Eventos

Instalaciones

Localizaciones

Organizaciones

Personas

Productos

Rostros

Koldo García	0:10 - 0:13	01:55 - 02:01
José Luis Ábalos	0:11 - 0:15	01:26 - 01:42
19	0:40 - 0:40	13:43 - 13:50
Ana Baldoín	01:13 - 01:18	06:30 - 06:33

AI is currently in production at the RTVE Archive

Audio and video analysis

- Automatic Speech transcription
- Speaker segmentation
- Keywords, entities and content classification
- Facial recognition
- Scene identification
- Object and logo recognition
- OCR

Multimodal analysis made by generative AI

- Automatic summarization
- Semantic segmentation

RTVE University of Zaragoza

Chair

- Created in 2017
- Our areas of collaboration include:
 - Benchmarking AI solutions in the market
 - AI training for the Archive and technical areas.
 - Identifying and developing quality criteria for tenders.
 - Promoting research on speech technologies by creating datasets for algorithm training.

RTVE Databases

- The RTVE Database is a collection of Spanish audio and video recordings from RTVE.
- It includes TV shows broadcast from 2015 to the present.
- The database is divided into partitions for training, development, and testing purposes.
- The main goal of the RTVE Database is to support research in speech recognition, speaker diarization, and multimodal diarization.
- These datasets have been used for the RTVE Albayzin challenges in Spanish.

<https://catedrartve.unizar.es/rtvedatabase.html>

Multimodal retrieval: a PoC

- Traditional Information Retrieval Systems (IRS) are typically based on full-text indexing.
- These systems rely on text or metadata describing the data in the audio or video files.
 - Human description
 - AI generated metadata
- This Proof of Concept (PoC) focuses exclusively on video content, without any extra metadata.
- This raises the critical question: how do we represent video-only content with no metadata?



Video encoder [0.1, 1.1, 0.2, ..., 0.7]



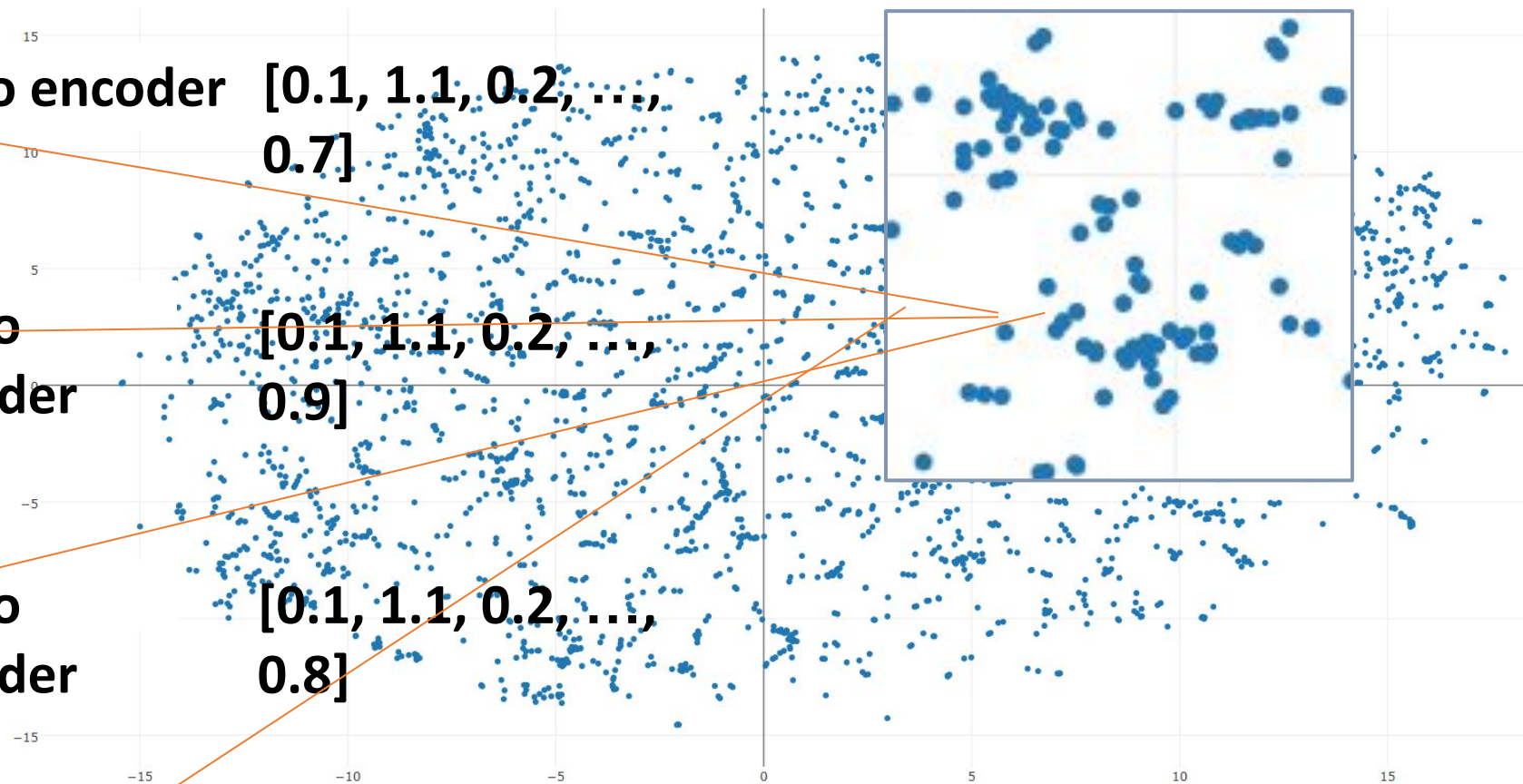
Video encoder [0.1, 1.1, 0.2, ..., 0.9]



Video encoder [0.1, 1.1, 0.2, ..., 0.8]

woman working in a factory

Text encoder [0.2, 1.0, 0.3, ..., 0.9]



Multimodal retrieval: research queries

- How can we explore our news silent film collections for production needs?
- How good is the performance of multimodal information retrieval technology?
- As the archive management system is not ready for this technology, is it worth it to integrate this metadata into the archive system? If so, how can this be done?
- Is this the new technological scenario for future information retrieval at Media Archives?

<http://signal4.cps.unizar.es:8502/>

A RTVE – University of Zaragoza PoC

rtve

Prueba de concepto búsqueda semántica multimodal

10 de junio 2024

Introducción

RTVE preserva un archivo filmico producido por TVE entre los años 1950 y finales de los 1980. Está formado por más de 170.000 latas que contienen programas, series y noticias. El objetivo principal, tras la digitalización de estas colecciones es hacer que el contenido sea accesible.

Una parte de la colección de noticias filmadas no tiene sonido y presenta títulos poco informativos como "Noticia número 5", sin embargo, contiene imágenes que son relevantes para ilustrar la evolución de la sociedad española. En este contexto, en colaboración con la Universidad de Zaragoza, a través de la Cátedra RTVE – Unizar, estamos explorando las posibilidades de los espacios semánticos para la representación y recuperación de información multimodal que no ha sido previamente **metadada**.

¿En qué consiste la búsqueda basada en espacios semánticos?

La recuperación de contenidos se realiza mediante una búsqueda en un espacio semántico multimodal que integra texto e imagen. Este espacio proporciona una manera estructurada y matemática de capturar y analizar el significado y las relaciones entre textos e imágenes. Tanto las imágenes como los textos se codifican en vectores multidimensionales, de modo que elementos semánticamente similares, como "gato", "perro" y una fotografía de un pastor alemán, estarán próximos en este espacio semántico debido a que representan conceptos similares, como animales domésticos.

Todo el contenido visual se segmenta automáticamente en planos con fotogramas semánticamente relacionados. Cada plano se representa mediante un vector en el espacio semántico y se almacena en una base de datos.

Cuando se realiza una búsqueda en lenguaje natural esta descripción se convierte en un vector semántico y se compara con todos los vectores semánticos de los planos almacenados en la base de datos. Los vectores más cercanos en el espacio semántico se recuperan y se muestran al usuario. Aunque la búsqueda semántica no está diseñada específicamente para localizar personas, monumentos, objetos u otros elementos concretos, algunos de estos pueden estar representados en el espacio semántico y, por lo tanto, pueden ser recuperados por el sistema, aunque no sea su objetivo principal.

¿Cuál es el objetivo de la prueba de concepto?

El objetivo es responder a estas preguntas de investigación:

- ¿Cómo podemos explorar nuestras colecciones de noticias filmadas sin sonido para las necesidades de producción?
- ¿Cuál es el potencial de la búsqueda semántica?
- Dado que el sistema de archivo no está listo para esta tecnología, ¿cómo podemos integrar estos metadatos en el sistema de archivo? Si es así, ¿cómo se puede hacer?
- ¿Es este el nuevo escenario tecnológico para la futura recuperación en el Archivo RTVE?

¿Qué contiene el conjunto de datos sobre el que se va a hacer la prueba?

200 contenidos procedentes del archivo de noticias filmadas de los años 1962 y 1963. Rodadas en 16mm BN sin audio y sin metadatos descriptivos en ARCA. Los contenidos seleccionados están rodados en exteriores y presentan variedad de planos y acciones. Cada contenido cuenta en la base de datos con una o dos frases breves a modo de descripción y un título.

169 contenidos procedentes del archivo de noticias filmadas de los años 1962 y 1963. Rodadas en 16mm BN sin audio. Los contenidos seleccionados están rodados en exteriores y presentan variedad de planos y acciones. Esta colección incluye **minutados** realizados por documentalistas expertos. Se comprueba que los minutos responden a una descripción lo más neutra posible de la imagen, es decir, que no cuentan con información que pueda deducirse de otras fuentes como el audio o el conocimiento experto del documentalista.

rtve

¿Cómo se han analizado los datos?

Se ha empleado un enfoque multimodal que integra texto e imagen para la representación y recuperación de información. Para ello, se ha segmentado automáticamente el contenido visual en planos con fotogramas semánticamente relacionados, utilizando CLIP de OpenAI para codificar las descripciones en lenguaje natural de los contenidos y las imágenes en vectores semánticos de 768 dimensiones. Con esta metodología, se ha logrado un total de 8.861 planos almacenados en la base de datos. Estos vectores, junto con las descripciones en lenguaje natural de los contenidos, se han utilizado para realizar búsquedas semánticas que permiten recuperar los elementos más relevantes.

Además, se ha trabajado en el desarrollo de una interfaz de consulta que facilita a los usuarios la búsqueda de contenido mediante consultas textuales o por imágenes y seleccionar los contenidos más relevantes para una búsqueda dada. La interfaz permite refinar la búsqueda a partir de la selección de uno o varios planos seleccionados por el usuario (refinar búsqueda). En el caso de consultas textuales, el texto escrito en español se traduce automáticamente a inglés. La figura 1 muestra la interfaz de consulta, donde básicamente existen dos campos para introducir la consulta textual o una imagen. El botón buscar inicia el proceso de búsqueda. Así, por ejemplo, si introducimos como consulta el texto "Rey Alfonso XIII", esta consulta se traduce a inglés como "King Alfonso XIII" y devuelve, ver figura 2, los planos semánticamente más parecidos a la consulta. En este caso nos devuelve 9 de 10 planos que contienen al Rey Alfonso XIII.

BuscaEscenas: rtvePoCArchivo



Figura 1. Interfaz de búsqueda de escenas

rtve

BuscaEscenas: rtvePoCArchivo

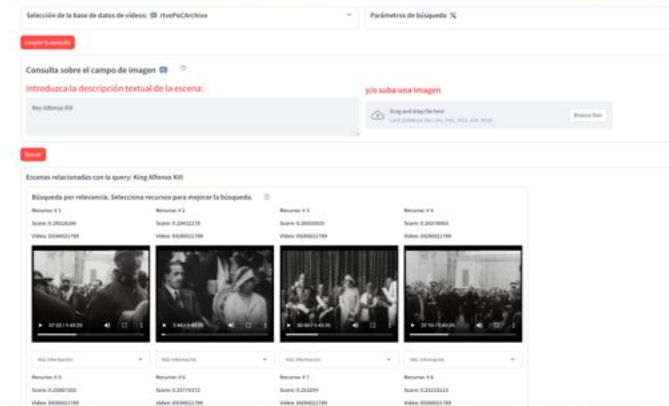


Figura 2. Resultados de la búsqueda "Rey Alfonso XIII". Se muestran los planos que el sistema de recuperación considera que están semánticamente relacionados con la consulta.

Objetivos de la prueba

El objetivo de la prueba es evaluar cómo funciona el sistema a partir de las búsquedas de los documentalistas. Para ello, se ha creado una interfaz muy simple donde escribir el texto de búsqueda o subir una imagen. Después de hacer clic en "Buscar", aparecerán los 10 resultados más relevantes. A la derecha, se mostrarán 10 casillas, una para cada resultado, y se debe marcar las que consideren correctas para la búsqueda realizada. Una vez marcadas, deben hacer clic en "Enviar Evaluación" para enviar su selección al servidor y obtener estadísticas sobre el funcionamiento del sistema. El sistema no guarda información personal y pueden limpiar la consulta para realizar una nueva búsqueda.

BuscaEscenas: rtvePoCArchivo

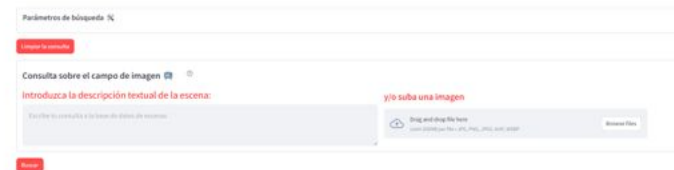


Figura 3. Interfaz de inicio para la evaluación del sistema de recuperación de contenidos.

Multimodal retrieval: Phases

Phase I

- Collection of 369 news items filmed between 1962 and 1963, without audio or previous descriptive metadata.
- They were filmed on location and present a variety of shots and actions.
- For two weeks the Archive team tested the system by performing natural language queries.
- Initial results were positive:
 - 87% of the 307 queries returned at least one relevant result in the top 10.

Phase II

- The data set was increased with 1,354 news items filmed in 1964 and 1965, which represents 27 more hours of video.
- We want to test if the system performance is just as good or even better.

Let's test the system performance

You are invited to search and try! The system supports queries both in English and Spanish

- Women and society
- Rural world
- School
- Family and domestic life
- Industry and economic activities

<http://signal4.cps.unizar.es:8502>

/

Collaborating with researchers helps us

- Identify the archive goals and priorities without the pressure of a supplier.
- Define relevant use cases as a way to reach short-term goals and show the relevance of embracing AI.
- Train archivists and adapt their expectations to the real performance of the technology.
- Be familiar with the latest developments in technology.
- As a public Archive, to support research in AI at university.

rtve.es

tve

tve 2

16:51:13