

PLIEGO DE PRESCRIPCIONES TÉCNICAS QUE RIGE LA CONTRATACIÓN DE LA AMPLIACIÓN Y ACTUALIZACIÓN DE LA PLATAFORMA ETIQMEDIA

- EXPEDIENTE NÚM. 2603OB04 -

1. DESCRIPCIÓN DEL SUMINISTRO

La CCMA, S. A. requiere ampliar y actualizar la plataforma del servicio de catalogación ya disponible desde el año 2021. Esta plataforma actualmente es la solución utilizada a nivel corporativo y está alojada en las instalaciones de la CCMA, S. A. con las funcionalidades descritas en el punto 2 y con su correspondiente servicio de mantenimiento asociado descrito en el punto 7.

Esta plataforma está compuesta por el software de Etiquedia en régimen de licencia perpetua y por el hardware necesario para poder procesar vídeos y audios con una volumetría estimada de 60h/día con las funcionalidades incluidas en la licencia así como la posibilidad de realizar entrenamientos de idioma para las transcripciones periódicamente.

Dada la utilización de la plataforma actual así como una previsión de crecimiento del volumen de horas a catalogar, se precisa hacer una ampliación de la arquitectura actual (ver punto 4).

2. DESCRIPCIÓN FUNCIONALIDADES INCLUIDAS EN LA LICENCIA

Módulos actuales:

- Módulo de corte e identificación de planos y escenas
- Módulo de reconocimiento de protagonistas/personas
 - Con herramienta de entrenamiento de manera autónoma por la CCMA, S. A.
- Módulo de lectura de textos en imágenes/vídeos (OCR)
- Módulo de reconocimiento de audio para los idiomas de catalán y castellano o castellana. En este módulo de manera excepcional se permitirá la utilización del servicio de *Google* para hacer la transcripción y traducción al catalán y, por lo tanto, cuando se utilice este módulo será el único momento donde la plataforma podrá extraer estos audios de la plataforma instalada en la CCMA, S. A.
- Módulo análisis semántico y categorización
- Módulo Tesauro con capacidad de Integración Tesauro
- Módulo Segmentación de voz y música
- Módulo Segmentación hablantes
- Módulo sincronización de guiones
- Herramienta gestión de la catalogación, vía aplicación HTML o escritorio, que permita las siguientes funcionalidades de:
 - Buscador de assets
 - Visualización en qué estado se encuentra cada paquete de trabajo y estado correspondiente del asset asociado (pendiente de procesar, procesado, en error, etc.)
 - Visualización de la metainformación generada por cada asset (transcripción, identificación personas, segmentación hablantes, etiquetas, tesauros...)

- Posibilidad de descargar en word, SRT y PDF la metainformación generada y asociada a cada *asset* procesado.
- Absorción texto con vídeo/audio, es decir, al posicionarse en un punto del texto que vaya al TC del *asset* y viceversa (bidireccional).
- Interfaz de entrenamiento de personas para su identificación con el módulo pertinente (introducción de nuevas personas, visualización de las personas ya introducidas, buscador de personas).
- API indexación:
 - El módulo que permite jerarquizar la orden de procesamiento, visibilizar el estado de la catalogación y aumentar la eficiencia en la recuperación de información desde el archivo.
- API Sincronización:
 - Implementación de un módulo de integración que conecta el sistema de envío de subtítulos con la plataforma de sincronización. Esta API permite gestionar prioridades en las tareas y monitorear en tiempo real el progreso de los trabajos, un factor crítico en la gestión de contenidos de larga duración como las películas.
- Resumen y titulares de noticias e identificación de secciones con IA Generativa:
 - Inteligencia artificial generativa para automatizar la segmentación semántica de los vídeos (identificación del inicio y fin de secciones) y la creación de resúmenes y titulares. Esta funcionalidad sustituye tareas que actualmente se realizan de forma manual, agilizando el trabajo de los documentalistas y facilitando el acceso rápido de la redacción al contenido indexado.

3. ARQUITECTURA ACTUAL

3.1 Arquitectura actual

La herramienta de gestión de catalogación actualmente funciona mientras la plataforma de indexación está trabajando y soporta 4 accesos concurrentes no nominales, es decir, desde 4 puestos de trabajo de manera concurrente en cualquier momento.

La plataforma, hoy en día, está dimensionada para poder realizar un procesamiento de 60h/día con la disposición de dos máquinas configuradas en modalidad de balanceo de carga.

Además se dispone de una tercera máquina que aloja el entrenamiento del modelo.

Estas 3 máquinas están instaladas en las infraestructuras de la CCMA, S. A. del campus Sant Joan Despí y se están gestionando de manera remota. Disponemos de una cuarta máquina para el entorno de PRE-Producción. Es física y totalmente autónoma de la arquitectura de Producción

Desde un punto de vista de integración de esta plataforma con los flujos, procesos y software de la CCMA, S. A. se utilizan las siguientes funcionalidades:

- El procesamiento de los archivos de media (vídeo y audio) se hace vía:
 - API
 - *Watchfolder* dentro de las máquinas de procesamiento
- El uso de la API permite:
 - Enviar *jobs* con detalle de qué trabajo o qué módulos debe realizar para el medio indicado.
 - De manera selectiva poder pedir (sólo etiquetas, sólo subtitulación...)
 - Para obtener status de un job
 - Reintentos en caso de errores (!)
 - Recoger metainformación generada
- La salida puede ser un único fichero, N ficheros por funcionalidad en formato JSON o XML. Por ejemplo, la obtención de la subtitulación directamente en formato SRT, WEBVTT, TTAF, si procede. O toda la información asociada a un asset directamente en un fichero de tipo JSON para a partir de esta salida realizar las tareas más adecuadas de posprocesar e integraciones con los sistemas o procesos corporativos.
- Además de la salida vía API y fichero(s) con formato backend (JSON,XML) para trabajar a nivel operatividad o de integración, también se dispone en la salida de los ficheros de transcripción detallados en diferentes formatos (word, pdf, txt) que se hacen llegar al solicitante del trabajo vía opción integrada dentro del MAM Corporativo.
- Perfiles de catalogación con diferentes contextos (idioma, términos, personajes, imágenes...)
- La arquitectura permite trabajar en concurrencia el procesamiento de material por cualquiera de las dos vías indicadas (API/*watchfolder*) y el acceso a la herramienta de catalogación del material para hacer búsquedas, visualizar los metadatos generados por un *asset*, validaciones, etc.
- En ningún caso el contenido enviado a esta plataforma sale de las máquinas dedicadas a este propósito con la excepción de los audios que enviamos al módulo de transcripción y traducción de Google por ser contenido con idiomas diferentes del castellano y catalán.

3.2 Especificaciones Hardware y administración de la plataforma actual

A continuación se hace una descripción de la composición y especificaciones de las máquinas de la arquitectura actual dentro de la plataforma alojadas en el CPD de TV3:

ENTORN PRO



ENTORN PRE



2 Máquinas de procesamiento de contenido audiovisual (vídeos y audios):

- Equipos enrackables en racks estándar de 19", con una profundidad de 90 cm, incluyendo guías y todo el material necesario para su instalación.
- Espacio máximo de 2U para cada equipo.
- Equipos independientes y autosuficientes. No se desean propuestas que engloben múltiples servidores en un chasis, con componentes compartidos
- Los equipos deben estar certificado para ejecutar:
 - Windows 2019 Server.
 - El fabricante proporciona los softwares necesarios para la instalación del sistema operativo anterior
- 1 Procesador Intel Xeon Silver 4214R o superior.
- Mínimo 64 GB de RAM.
- Controladora SAS/SATA con capacidad de RAID 0,1,5. El RAID debe hacerse por hardware, sin necesitar implicación ni drivers instalados en el SO. Hardware para caché con un mínimo de 1GB de RAM. Caché de escritura protegido por batería/supercapacitador.
- Capacidad de soportar como mínimo 6 discos SAS/SATA Hotplug de 2,5"
- 2 discos de estado sólido SATA de 1TB hot swap
- Tarjeta gráfica NVidia Quadro RTX 4000 de 8GB
- 4 puertos de red ethernet 1000-BASE-T en placa base o tarjeta PCI.
- Alimentación eléctrica y refrigeración
 - Fuentes de alimentación redundantes y sustituibles en caliente, nivel de eficiencia Platinum.
 - Ventiladores para refrigeración redundantes
- El equipo debe incorporar un procesador de servicio que permita la gestión remota, incluyendo soporte para IPMI 2.0 y el acceso a la consola, en modo gráfico, a través de la red IP (KVM virtual) mediante un puerto de red dedicado.
- 2 HDD mecánico de 6 TB hot swap

1 Máquina para alojar el modelo reducido:

- Equipo enrackable en racks estándar de 19", con una profundidad de 90 cm, incluyendo guías y todo el material necesario para su instalación.

- Espacio máximo de 2U para cada equipo.
- Equipo independiente y autosuficiente. No se desean propuestas que engloben múltiples servidores en un chasis, con componentes compartidos
- El equipo debe estar certificado para ejecutar y debe incorporar licencia para:
 - Windows 10
 - Ubuntu 20.04.2 LTS
 - El fabricante proporciona los softwares necesarios para la instalación del sistema operativo anterior
- Procesador Intel Core i7-9700,(8 Core, 12MB Caché, 3.0Ghz) o superior.
- Mínimo 64 GB de RAM.
- Tarjeta gráfica NVidia RTX3080.
- 2 puertos de red ethernet 1000-BASE-T en placa base o tarjeta PCI
- Controladora SAS/SATA con capacidad de RAID 0,1,5. El RAID debe hacerse por hardware, sin necesitar implicación ni drivers instalados en el SO. Hardware para caché con un mínimo de 1GB de RAM.
- 2 discos de estado sólido SATA de 1TB hot swap
- 2 HDD mecánico de 6 TB hot swap
- Alimentación eléctrica y refrigeración
 - Fuentes de alimentación redundantes y sustituibles en caliente, nivel de eficiencia Platinum.
 - Ventiladores para refrigeración redundantes

1 Máquina de procesamiento de contenido audiovisual – Entorno Preproducción (vídeos y audios)

- Equipo enrackable en racks estándar de 19", con una profundidad de 90 cm, incluyendo guías y todo el material necesario para su instalación.
- Espacio máximo de 2U para cada equipo.
- Equipo independiente y autosuficiente. No se desean propuestas que engloben múltiples servidores en un chasis, con componentes compartidos
- El equipo debe estar certificado para ejecutar y debe incorporar licencia para:
 - Windows Server 2022
 - El fabricante proporciona los softwares necesarios para la instalación del sistema operativo anterior
- Procesador Intel Core i7-9700,(8 Core, 12MB Caché, 3.0Ghz) o superior.
- 32 GB de RAM.
- Tarjeta gráfica Nvidia RTX4000 de 16 GB
- 2 puertos de red ethernet 1000-BASE-T en placa base o tarjeta PCI
- Controladora SAS/SATA con capacidad de RAID 0,1,5. El RAID debe hacerse por hardware, sin necesitar implicación ni drivers instalados en el SO. Hardware para caché con un mínimo de 1GB de RAM.
- 2 discos de estado sólido SATA de 1TB hot swap
- 2 HDD mecánico de 6 TB hot swap
- Alimentación eléctrica y refrigeración
 1. Fuentes de alimentación redundantes y sustituibles en caliente, nivel de eficiencia Platinum.
 2. Ventiladores para refrigeración redundantes

El acceso para administrar estas máquinas se hace vía VPN de la CCMA, S. A.

4. PROPUESTA ARQUITECTURA AMPLIADA

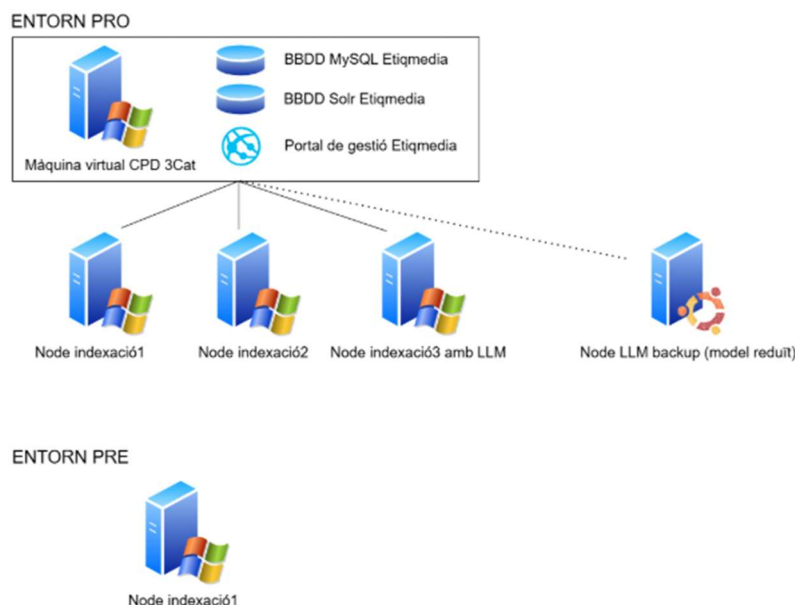
La arquitectura actual, dimensionada para una capacidad máxima de 60 horas diarias, se ha convertido en insuficiente por dos motivos principales:

1. Incremento de Volumen de Procesamiento: La previsión de carga para los próximos años se estima en un mínimo de 75 horas diarias, superando la capacidad máxima de la configuración actual y requiriendo un nodo de procesamiento adicional.
2. Incorporación de Funcionalidades Críticas (LLM): Los nuevos módulos requeridos por la CCMA, S. A. (Título, Resumen, Corte Semántico - ver Punto 2) requieren el uso intensivo de Modelos de Lenguaje (LLM) con requerimientos de tarjeta gráfica (GPU) de alto rendimiento. La arquitectura actual no permite su incorporación óptima ni garantiza el rendimiento y la estabilidad exigidos para la volumetría de 75 horas/día.

Por estos motivos, se hace necesaria la propuesta de Arquitectura Ampliada y Actualizada detallada a continuación.

4.1 Arquitectura ampliada

Se propone la actualización de la arquitectura de la plataforma, para poder añadir un nodo adicional de indexación que permita aumentar la capacidad de procesamiento, pero también para independizar las BBDD, el portal y la API de los nodos de indexación:



Se propone instalar en una máquina virtual Windows Server la BBDD del Solr 9, la BBDD MySQL, la API y el portal de gestión de la plataforma. Esta máquina supone el elemento central del sistema y utilizar un entorno virtual con los backups asociados facilita una recuperación muy rápida ante cualquier problema.

Esta máquina virtual debería tener las siguientes características:

- CPU: 8 Cores 2.4Ghz
- Memoria: 64 GB RAM
- Disco duro SO y base de datos: 1 TB HD (SSD Recomendado)

Además de la máquina virtual, se contaría con tres nodos de procesamiento, capaces de procesar como máximo 30 horas diarias cada uno para todos los algoritmos, excepto para los que requieran uso intenso de LLM (p. ej.: resúmenes y titulares).

Para el caso del LLM se propone instalar el LLM en uno de los nodos, y desde los otros dos hacer uso de este LLM para realizar las tareas de resumen. Esto nos permitiría 30 horas de resúmenes diarios.

Para este nodo, que tiene instalado el modelo de lenguaje (LLM), se requerirá una tarjeta gráfica más potente (con al menos 24 GB de memoria) para poder cargar el modelo LLM en su totalidad para acceso rápido y más eficiente.

Dada la criticidad que supone disponer únicamente de un nodo que tenga ubicado el modelo, se propone tener un nodo LLM de backup en la actual máquina de entrenamiento, que utilice una versión del LLM cuantificada que quepa en la tarjeta actual.

Con esta arquitectura podríamos reutilizar las máquinas actuales como nodo de procesamiento 1 y nodo de procesamiento 2, y se propone la incorporación de una nueva máquina como nodo de procesamiento 3 que tendrá las siguientes características:

- CPU: 12 Cores 2.4Ghz
- Memoria: 32 GB RAM
- Disco duro SO y Base de Datos: 1 TB HD (SSD Recomendado)
- Disco duro Media: 2TB HD (Mínimo según necesidad de almacenamiento)
- 1 Tarjeta gráfica: Nvidia L4 24GB o similar.
- 1 tarjeta Nvidia RTX4500 16GB o similar.

5. REQUERIMIENTOS DE LICENCIA DE LA PLATAFORMA ETIQMEDIA

El adjudicatario debe incluir en su oferta el coste de la ampliación y la provisión de licencias de software necesarias para habilitar el paso de la capacidad de procesamiento de 60 horas diarias a un mínimo de 75 horas diarias mediante la nueva arquitectura propuesta.

Además, se requiere la inclusión del coste para la adquisición de nuevas funcionalidades previstas para el desarrollo futuro.

5.1. Ampliación de Licencias de Módulos Existentes (Incremento de 15 horas/día)

Con el fin de habilitar la nueva capacidad de procesamiento de hasta 75 horas diarias, se requiere la ampliación de licencia para los siguientes módulos existentes en un tramo adicional de 15 horas/día por encima de la licencia base de 60 horas:

Módulo (Funcionalidad)	Tramo de Ampliación Requerido
Módulo de Detección y Segmentación de Hablantes	15 horas adicionales

Módulo de Reconocimiento del Idioma del Audio	15 horas adicionales
Módulo de Traducción del Catalán o Castellano	15 horas adicionales
Módulo de Análisis Semántico y Categorización	15 horas adicionales
Módulo Detección y Segmentación voz, música y silencios	15 horas adicionales
Módulo Medición Calidad de audio y selección de pista	15 horas adicionales

5.2. Adquisición de Nuevas Funcionalidades

La oferta debe incluir, también, la adquisición de la licencia y la preparación técnica para las nuevas funcionalidades de investigación avanzada que se prevé incorporar a la plataforma durante el año 2027.

Se propone la opción de evolucionar el sistema con nuevas funcionalidades como una ampliación posterior. Esta adquisición funcional se haría como una segunda fase una vez esté estable la ampliación de arquitectura y funcionalidad explicada en el punto 4. Se propone principalmente el trabajo con las búsquedas semánticas (de texto y de imagen) y la búsqueda biométrica.

Los módulos que se quieren adquirir son:

1. **Módulo de búsqueda semántica textual:** Permite hacer búsquedas en las transcripciones de contenido utilizando lenguaje natural, sin depender de palabras clave literales. El sistema utiliza modelos *transformer* y *embeddings* para encontrar los resultados más cercanos (mediante la distancia coseno). Incluye la posibilidad de filtrar por entidades y devuelve información detallada como títulos, fragmentos relevantes, puntuación y fechas.
2. **Módulo de búsqueda semántica de imagen:** Este módulo permite encontrar fotografías o planos de vídeo a partir de una descripción textual del usuario. En lugar de comparar etiquetas o palabras clave, el sistema "comprende" lo que se describe gracias al uso de *embeddings* semánticos que vinculan texto e imagen, localizando visualmente el contenido que mejor encaja con la petición.
3. **Módulo de búsqueda biométrica de imagen:** Permite localizar personas en archivos de vídeo o fotos introduciendo simplemente una imagen de su cara. La gran ventaja es que no es necesario que la persona esté previamente identificada por su nombre en el sistema. Utiliza *embeddings* faciales (descripciones matemáticas del rostro) para encontrar apariciones de cualquier protagonista, aunque no hubiera sido considerado relevante en el momento de incorporar el contenido en la base de datos.

Se requiere la adquisición de los siguientes módulos para un total de 75h/día:

Módulo (Funcionalidad)	Tramo de Ampliación Requerido
Módulo de búsqueda semántica textual	75 horas adicionales
Módulo de búsqueda biométrica de imagen	75 horas adicionales
Módulo de búsqueda biométrica de imagen	75 horas adicionales

6. FASE INSTALACIÓN Y CONFIGURACIÓN

La plataforma debe estar instalada y configurada en un máximo 3 meses después de la formalización del contrato en las instalaciones de la CCMA, S. A.

La instalación mínima requerida una vez transcurridos estos 3 meses y para un procesamiento de 75h incluye:

1. Arquitectura definida en el punto 4
2. Ampliación de los módulos necesarios funcionalmente:
 - e. Módulo de Detección y Segmentación de Hablantes
 - f. Módulo de Reconocimiento del Idioma del Audio
 - g. Módulo de Traducción del Catalán o Castellano
 - h. Módulo de Análisis Semántico y Categorización
 - i. Módulo Detección y Segmentación voz, música y silencios
 - j. Módulo Medición Calidad de audio y selección de pista
 - i. Nueva versión ligera de la herramienta gestión de la catalogación

En esta fase también se incluye un mínimo de 3 sesiones de formación de 3h de duración para personal de la CCMA, S. A.

Una vez certificada la instalación y configuración por parte de la CCMA, S. A. para la capacidad de procesamiento de 75h/por nodo con los módulos descritos se procederá al pago según el precio de licencia de la licitación.

7. CONTRATO DE ACTUALIZACIÓN, MANTENIMIENTO Y SOPORTE

La oferta debe incluir una propuesta de mantenimiento del software con las actualizaciones de mejora de los módulos adquiridos, y los servicios y tareas necesarios para mantener su buen funcionamiento que se describen a continuación.

En concreto, el soporte requerido incluye:

- Actualización de versiones de la plataforma
- Apoyo técnico en caso de incidencia con este acuerdo de nivel de servicio:
 - Soporte 12x5
 - Tiempo de respuesta a incidencia: 24h máximo
 - Tiempo de resolución a incidencia: 48h máximo
- Tareas incluidas en el mantenimiento
 - Adaptación de los flujos de entrada y parametrización de la plataforma para nuevos tipos de contenido que requieran nuevos perfiles según el contexto del contenido
 - Formaciones en el caso de adquirir nuevos módulos o que una actualización de cualquier módulo instalado requiera una nueva formación para los cambios introducidos
 - Soporte para consultas o dudas relacionadas con el funcionamiento de la plataforma o para encontrar explicaciones a comportamientos anómalos.