
**GESTIÓN, EXPLOTACIÓN INDIZACIÓN DE
ONTOLOGÍA DE BUSCADOR DE RTVEPLAY**
PLIEGO DE CONDICIONES TÉCNICAS



Índice	Página
1 Introducción.....	3
2 Objeto del pliego de condiciones técnicas.....	3
3 Descripción del Servicio	3
4 Estructura solicitada para la propuesta técnica	6
Anexo 1.- Especificaciones de la API de Distribución de RTVE	7

1 Introducción

La introducción de las tecnologías de Inteligencia Artificial y procesamiento de lenguaje natural en los procesos productivos de RTVE es un elemento estratégico dentro de los planes tecnológicos con el propósito de mejorar los servicios que se ofrecen a los usuarios, así como simplificar y automatizar procesos que actualmente no se llevan a cabo en RTVE.

Dentro de los desarrollos realizado en el contexto del proyecto RTVEPlay cabe destacar en este sentido la implementación de un buscador semántico, que se apoya en una ontología creada con el propósito específico de optimizar las búsquedas de los programas, películas, documentales y contenidos audiovisuales en general disponibles a través del portal de rtve.es, y en particular en el entorno de Play.

La razón de haber creado una ontología para esta labor se justifica en la dificultad de realizar búsquedas sólo basadas en el contenido en bruto, dada la amplitud y diversidad del catálogo de RTVE.es, siendo necesario identificar entidades y abstraer aquella información que se considera que puede utilizarse para clasificar y relacionar los contenidos entre sí, además de disponer de mecanismos para explotarla, de cara a obtener resultados lo más ajustados posible en tiempos aceptables para un usuario de los servicios de Play.

Una vez definida la ontología, fue necesario implementarla en un motor para su explotación basado en técnicas de procesamiento de lenguaje natural (PLN). Tanto la ontología como las reglas de conocimiento definidas están disponibles en un formato estándar (OWL, JSON y CSV) compatible con software disponible en el mercado.

2 Objeto del pliego de condiciones técnicas

El objeto de este pliego es establecer las características técnicas imprescindibles que rigen la contratación de un servicio para la importación, gestión y mantenimiento de la ontología y reglas de conocimiento creadas para el buscador semántico de RTVEPlay. Será necesario gestionar su explotación sobre un motor de conocimiento asegurando su correcto funcionamiento, disponibilidad y rendimiento.

Además, es necesario adaptar el proceso de indización de contenidos con formato de RTVE para el motor de conocimiento que se proponga a utilizar.

3 Descripción del Servicio

Dada la naturaleza semántica del buscador de RTVEPlay, se requiere de un servicio que lleve a cabo la gestión de la ontología y reglas de conocimiento que se han creado para dicha labor sobre un motor de conocimiento que permita explotarla. Así, surgen diferentes necesidades asociadas al proyecto:

- Importación en el motor de conocimiento de la ontología definida en el formato estándar OWL y reglas de conocimiento que se han generado para el buscador de RTVEPlay (patrones de ortografía y clasificación gramatical para facilitar la tokenización e identificar acrónimos, fechas, números, excepciones, ...).
- Gestión y mantenimiento de la ontología y de las reglas de conocimiento existentes para asegurar su correcta explotación por el motor de conocimiento.
- Operación, gestión y mantenimiento del motor de conocimiento, ofreciendo el soporte necesario para asegurar que presta el servicio correctamente para el proceso de búsqueda.

En lo que se refiere a la importación, gestión y mantenimiento de la ontología y las reglas de conocimiento, RTVE valorará la propuesta del licitador del servicio en lo relativo a la correcta importación y gestión sobre el vocabulario controlado que se ha definido en el contexto del proyecto del buscador. En particular deberán poder importarse y gestionarse:

- Entidades identificadas en la ontología:
 - o Crear, modificar o borrar entidades.
 - o Gestión de entidades, añadiendo atributos, alias, etc.
 - o Control de elementos duplicados y detección de errores.
- Relaciones definidas en la ontología.
 - o Crear, modificar o borrar relaciones entre elementos.
 - o Control de posibles inconsistencias.
- Clústeres y clasificaciones:
 - o Posibilidad de crear, modificar o borrar elementos en los clústeres existentes.
 - o Mantenimiento de clústeres existentes y detección de inconsistencias.
- Reglas de conocimiento:
 - o Crear, modificar o borrar reglas para garantizar el correcto funcionamiento de procesos como la tokenización e identificación de números, fechas, etc.
 - o Control de duplicados y posibles inconsistencias.

Adicionalmente, el licitador deberá facilitar una consola de gestión que ofrezca un **interfaz amigable** sobre este vocabulario controlado, de cara a que pueda ser utilizado por personal designado por RTVE, y que permita:

- Cargar la ontología
- Realizar manualmente por parte de un operador las acciones detalladas con anterioridad de creación, modificación y borrado de aquellos elementos que RTVE decida gestionar.
- Posibilidad de definir perfiles de usuario para el control de acceso y modificación de la información.
- Persistir la ontología en forma de ficheros o en una base de datos estructurada.

Por otro lado, desde el punto de vista del motor de conocimiento para la explotación de la ontología gestionada y las reglas de conocimiento, es necesario que el licitador provea un sistema basado en tecnologías de procesamiento de lenguaje natural (PLN) que opere sobre lo definido previamente. Se debe presentar información sobre la solución propuesta, de cara a que pueda ser valorada por parte de RTVE desde los siguientes aspectos:

- Punto de vista técnico:
 - o El servicio debe prestarse desde infraestructura onpremise de RTVE, debiendo el licitador describir la arquitectura de la solución, teniendo en cuenta que:
 - Debe operar como un servicio web, con posibilidad de ofrecer servicio a través de servidores Apache, NGINX o IIS.
 - Debe permitir la redundancia del servicio para ofrecer alta disponibilidad. El servicio debe ofrecer una capacidad de respuesta de al menos 50 peticiones por segundo.
- Punto de vista funcional:
 - o Deberá detallarse en qué se basa el proceso de búsqueda de la solución y cómo se apoya en el motor de conocimiento, atendiendo a las búsquedas introducidas por los usuarios, tokenización de dichas búsquedas, proceso de clasificación gramatical, reconocimiento de entidades, identificación de relaciones entre los contenidos, detección de patrones de fechas y expresiones temporales, etc.
 - o Se valorarán detalles sobre tecnología utilizada, APIs suministradas, metodologías/algoritmos utilizados para procesamiento de lenguaje natural

(tokenización, detección de entidades, clasificación en clústeres semánticos, indexación de reglas, detección de patrones, metapropiedades, ...).

También, de cara a integrar la solución dentro del sistema del buscador de RTVEPlay, es necesario realizar las siguientes adaptaciones:

- Proceso de enriquecimiento de búsquedas:
 - Datos:
 - El API de búsqueda de RTVE, desarrollada en NodeJS, que utiliza como repositorio de datos tecnología Elasticsearch, donde se encuentran almacenados los objetos susceptibles de ser devueltos como resultado
 - El literal de búsqueda introducido por el usuario para el que se espera, una vez procesado por el motor de conocimiento, que se hayan identificado sinonimias, entidades, fechas o cualquier otro elemento que provea el motor de conocimiento.
 - Es necesario adaptar el proceso de construcción de queries sobre Elasticsearch del API de RTVE al formato de información devuelto por el motor de conocimiento suministrado en la solución para devolver resultados más allá de la mera búsqueda del literal inicial.
- Proceso para incorporar nuevo conocimiento a la base de datos de la ontología de referencia que se ha definido.
 - Es necesario disponer de capacidad de indización de contenidos. Para ello, habrá que adaptar el proceso de indización de contenidos actual al motor de explotación facilitado por el licitador, teniendo que procesarse contenidos con el formato API propio de RTVE para programas, vídeos y audios. Para más detalles consultar el anexo técnico.
 - La finalidad de este proceso es poder incorporar la información que contienen posibles nuevos contenidos o reprocesar los que ya se hubiesen incorporado a la base de datos previamente, identificando las entidades, estableciendo relaciones, incorporando los elementos a los clústeres y clasificaciones existentes, etc.
 - Para el facilitar la adaptación de este proceso se facilitará el código del proceso indizador que existe actualmente, pudiendo tomarse como base.

Se valorará propuesta del licitador para llevar a cabo esta adaptación, indicando qué infraestructura sería necesaria y tecnologías utilizadas.

Finalmente, además de lo anterior, el licitador tendrá la obligación de:

- Dar formación sobre las herramientas al personal de RTVE que se pudiese designar para operar sobre alguna parte del sistema.
- Instalación y puesta en marcha del software.
- Durante el periodo de prestación del servicio:
 - Aportar todas las licencias del software necesarias para la gestión y el correcto funcionamiento del buscador.
 - Mantenimiento de las herramientas de gestión.
 - Asistencia al servicio ante posibles incidencias.

4 Estructura solicitada para la propuesta técnica

Descripción detallada de la solución:

- Documentación técnica del servicio que se propone como solución:
 - En su caso, el detalle de las licencias software necesarias e incluidas en la propuesta del licitador, describiendo el modelo de licenciamiento en el que operarán de cara a identificar posibles limitaciones del servicio (servidor, peticiones, usuarios, ...).
 - Gestión de la ontología y reglas de conocimiento:
 - Detalle de las actividades a realizar para la gestión y mantenimiento de la ontología y las reglas de conocimiento.
 - Propuestas para posibles mejoras y evolución.
 - Consola de gestión del servicio: interfaces disponibles (online vs escritorio), gestión de perfiles de usuario, posibilidades de importación/exportación de la información a formatos estándar, detalles del interfaz amigable solicitado, ...
 - Motor de conocimiento:
 - Detalle de arquitectura de la solución de cara a cumplir con los requisitos técnicos especificados.
 - Detalle funcional de proceso de búsqueda de la solución y tecnologías y algoritmos sobre las que se apoya.
- Planificación de puesta en marcha de sistema, junto con el detalle del proceso de importación y desarrollos para la adaptación de los procesos de búsqueda e indización al motor de conocimiento de la propuesta.
- Detalles de asistencia técnica al servicio:
 - Mecanismo de apertura de incidencias (correo, teléfono, herramienta de tickets, ...)
 - Tipo de asistencia que se facilitará (5x8, 24x7, ...)
- Plan de formación para personal designado de RTVE.

Madrid, noviembre de 2022

Anexo 1.- Especificaciones de la API de Distribución de RTVE