

Nº Expte.: 183028-18

OFICINA TÉCNICA DE PARTICIPACIÓN,
TRANSPARENCIA Y GOBIERNO ABIERTO

- **Punto 4.1: Análisis del proceso de crawling.**

- *Su proceso de indexación. ¿hace directamente crawling de las distintas web del ayuntamiento para obtener el contenido a indexar, o consulta las bases de datos donde residen esos contenidos, para obtener la información a indexar?*

Hace el crawling a través de los enlaces existentes en la sede electrónica.

- *Se indica que habrá que indexar el número de visitas de cada página, y que esta info la proporcionará el departamento de IT. ¿Cómo la proporcionará, existirá una BD a la que poder consultar el número de visitas asociadas a un contenido?*

Los ficheros de registro de accesos por http estarán alojados en la plataforma Elastic search.

- *Se indica que habrá que indexar también para cada contenido si su versión móvil es válida, y que esta información se obtendrá de una fuente de información seleccionada por el Ayuntamiento. Necesitaríamos más detalles de esto. ¿Habrá que integrarse con alguna aplicación o web que reciba el contenido e indique si es válido o no para móvil? ¿O ésta información estará precargada en una BD u otra fuente, de forma similar al número de visitas de cada página?*

Esta fuente se seleccionará de algún servicio externo que disponga de una API para solicitar esta información.

- **Punto 4.3: Migración a Elasticsearch del motor de búsqueda de www.zaragoza.es**

- *Este punto está relacionado con el 4.2, en el 4.2 se crea una guía de migración y en el 4.3 se da soporte al personal municipal para que, utilizando la guía, migren el motor de búsqueda de la web. ¿Es correcto?*

Así es

- **Punto 4.3: Indexación de los contenidos de la hemeroteca de Zaragoza.**

- *Indexación de los fondos digitalizados de la hemeroteca. ¿En qué formato se encuentra esta información? ¿Son imágenes que habrá que tratar con alguna aplicación para extraer el texto a indexar? ¿Están almacenados en una BD junto con metainformación (origen, tipo, título, autor, etc.) que podamos utilizar para indexar*

Se trata de 750.000 ficheros que se encuentran en formato tiff grupo 6 color y pdf/A con ocr incluido.

El texto a indexar está en ficheros con ALTO OCR en formato .xml

La metainformación necesaria para indexar se obtendrá a partir del nombre del fichero, que permite la asociación con el título de la publicación periódica y con la fecha:

Ejemplo: HMZ_0016_1955_12_02.pdf = El Noticiero de 2 de diciembre de 1955

Donde: HMZ = Hemeroteca Municipal de Zaragoza

0016 = Nombre de la publicación (se proporcionará un listado con todos los códigos y sus correspondientes nombres)

1955_12_02 = Año, mes y día

- **Punto 5.1: Adaptación a las directrices de diseño**

- *El pliego incluye solamente la migración del motor de indexación y la indexación de los fondos digitalizados, en ambos casos sin interfaz gráfica asociada, por lo que esta guía de estilos no aplicaría.*

Así es

- **Punto 5.4: Entorno tecnológico**

- *Habla de java 1.6, pero tanto Elasticsearch como sus librerías cliente requieren java 1.8, ¿es correcto?*

Las aplicaciones cliente se adaptarán para poder utilizar la tecnología ELK con Java 8

- *Aparte, necesitaríamos saber el volumen aproximado de información a indexar (número de contenidos, número de documentos digitalizados, etc.)*

En la sede electrónica existen alrededor de 250.000 contenidos

- *Respecto al hosting, entendemos que es tarea del Ayuntamiento, así como el montaje de Elasticsearch en clúster, etc. es decir, no habría que incluir en la oferta nada relacionado con la configuración de los servidores, configuración de Elasticsearch, etc., ¿correcto?*

La plataforma ElasticSearch estará operativa sobre los servidores corporativos, este proyecto sólo afecta a las configuraciones necesarias para indexar los contenidos referidos en el punto 4.

I.C. de Zaragoza a 21 de noviembre de 2018

El Jefe de la Unidad de
Gestión de la Web Municipal



Fdo.: Víctor Morlán Plo

La jefa de la Oficina Técnica de
Participación, Transparencia y
Gobierno Abierto,



Fdo.: Mª Jesús Fernández Ruíz