



Universidad Central de Venezuela.
Facultad de ciencias.
Postgrado Ciencias de la Computación.

Curso. Análisis de Documentos en Archivos Web.
Profesor. Andres Sanoja
Autor. Erikson Agustin Rodriguez Morillo.

Solución web de consulta de blockset utilizando api de BOM-o-MATIC “SEGQL”

Resumen

Desde el inicio y de igual manera en la actualidad el mundo de la computación gana valor por la velocidad en análisis de la información a gran escala, manejo de grandes volúmenes de datos, incluso forma parte fundamental de resultados y procesos de los sistemas de información. La web es uno de esos ejemplos de densidad de información. En efecto, al momento de la creación de la web y con la popularidad de integración entre tecnologías ha crecido de manera exponencial la información que se ha cargado y se encuentra disponible en ella. En consecuencia se ha incrementado la necesidad de búsqueda y organización de este gran almacén de datos. Con intención de hacer búsquedas, obtener características de datos y diseño en estructuras, han surgido celosos proyectos que manejan estos procesos. Block-o-Matic es un proyecto que analiza documentos web mediante métricas, así se obtienen un conjunto de bloques según los parámetros de configuración establecidos. Estos bloques contienen la información de las regiones en el documento web analizado. Este proyecto posee un API para su utilización, partiendo de esta se crea como proyecto hijo esta solución. La solución tiene el objetivo de listar en un browser bloques de una segmentación según las condiciones y ordenamiento establecidos en la consulta.

Desarrollo

Este proyecto consta de componentes Frontend, encargados de cargar la documentación del proyecto, dos componente de captura y envío de la consulta al servidor, por otro lado, y como respuesta a la solicitud se dispone de un componente desacoplado donde se carga el mensaje contenido, ya sea el mensaje informativo sobre algún error generado en la estructura de la consulta o el numero de columnas retornadas en la consulta.

Por ultimo, se implemento un componente dinámico presentado por medio de una tabla con el detalle de cada segmento retornado de acuerdo a los elementos descritos en la consulta.

Tecnologías

La solución esta implementada en el lenguaje de programación Ruby, utilizando un marco de trabajo amigable y muy productivo para el desarrollo web llamado Sinatra, esto brinda implementaciones de comunicación web.

Por otra parte para el mantenimiento de la aplicación o requerimiento adicional se implemento bundle para mantener versiones de tecnologías compatibles utilizadas a lo largo del proyecto.

El desarrollo se hizo a través de la utilización de herramientas que soportan la codificación del lenguaje, como editor se empleo Visual Studio Code, herramienta que mediante extensiones adicionales añade características que apoyan la productividad de desarrolladores. Así mismo, en el control de versiones se utilizo git para mantener los cambios y mejoras dentro del proyecto durante esta etapa piloto de la solución.

En cuanto a la tecnologías implementadas dentro del proyecto ademas de Ruby como lenguaje y Sinatra como marco de trabajo se vinculo las vistas a Bootstrap 4, para dar un estilo adecuado a la vista del usuario.

Como un estándar para las dependencias del proyecto en cuanto a las librerías utilizadas se hizo la instalación de Bundle, a través de la utilización del archivo Gemfile.

Alcance

Una vez analizado los recursos disponibles de acuerdo al entorno, se presentaron ciertos desafíos para la implementación de todas actividades que el proyecto requería. Aun de esta forma, se continuo el desarrollo y según la disposición de los recursos se desarrollo la solución, basado en el archivo segmentacion.json con el blockset generado por el api mientras se encontraba online, teniendo en cuenta que es un URL valido. Sin embargo no va directamente contra el repositorio del BOM. Considerando este punto, existe la limitación para esta versión del proyecto. De esta manera, se puede demostrar el manejo de la estructura de la consulta, así mismo con el conocimiento de los parámetros contenidos en el blockset, se pueden configurar criterios o condiciones y luego obtener las filas según criterio explícitamente configurado en ella.

Se tiene en cuenta que en un ambiente donde se disponen recursos mínimos necesarios, se podría incorporar la generación directa según el URL almacenado en el repositorio BOM.

Conclusión

Se presentan situaciones donde en ambientes desafiantes es necesario estudiar los recursos disponibles para encontrar una manera organizada de obtener los resultados que necesitamos. Este proyecto fue desarrollado tomando los recursos disponibles y moldeando los resultados para obtener una vista organizada de una solicitud. Durante el desarrollo se presentaron múltiples escenarios donde a través del análisis se construyo la solución a cada uno de los retos. De esta forma se avanza hacia el objetivo planteado en principio. Aun así, siempre se tuvo en cuenta un escenario común, donde el volumen de información y cantidad de variables fue el común denominador, por lo tanto fue considerado para el manejo de grandes cantidades de datos, en todo caso es necesario tener en cuenta el mayor numero de variables conocidas y mecanismos de normalización, y así obtener un resultado objetivo y claro.