

Trabajo Práctico Especial

- Consulta y

Transformación de

documentos XML

Grupo 7

Agustin Korman, Rodrigo Hernandez, Eduardo Tormakh

Fecha entrega: 12/11/2024

INTRODUCCIÓN

Este trabajo práctico tiene como objetivo aplicar herramientas avanzadas de consulta y transformación de documentos XML, con un enfoque en el uso de XQuery y XSLT. A través del proyecto, se utiliza la API de Congress.gov para obtener información detallada sobre el Congreso de los Estados Unidos, incluyendo datos sobre sus integrantes, cámaras y sesiones. La meta es automatizar el proceso de extracción y organización de estos datos, generando un archivo XML consolidado que, posteriormente, se transformará en una página HTML accesible y bien estructurada. Esta página presentará información clave del congreso, como los miembros de cada cámara, los períodos de las sesiones y otros detalles relevantes.

DESARROLLO

Al inicio del proyecto, el equipo decidió leer en conjunto la consigna para asegurarse de entender claramente los objetivos del trabajo y tener una visión global de las tareas a realizar. Después de eso, todos los integrantes visitaron la página de la API de Congress.gov, donde generaron sus propias API Keys para poder realizar consultas de manera individual. Con esta configuración completada, procedimos a ejecutar el archivo "tpe.sh", que automatiza el flujo de trabajo y genera cuatro archivos: "congress_info.xml", que contiene información general del congreso, como las sesiones y el periodo; "congress_members_info.xml", que incluye la lista de integrantes de cada cámara del congreso; "congress_data.xml", un archivo intermedio generado con "extract_congress_data.xq", que organiza los datos en un formato adecuado para la transformación final; y "congress_page.html", una página HTML generada mediante "generate_html.xsl" a partir de "congress_data.xml", presentando la información del congreso en un formato accesible y visual.

Para garantizar el manejo adecuado de errores, implementamos cuatro tipos de mensajes de error. El primero, 'download_member_error', se muestra cuando ocurre un problema en la descarga de "congress_members_info.xml", mientras que 'download_congress_error' indica problemas al descargar "congress_info.xml". También incluimos 'null_api_key', que se activa cuando la API Key no está configurada, impidiendo realizar las consultas, y 'invalid_argument', que verifica que el parámetro congress_number sea un número entre 1 y 118. En los casos en que no se encuentran registros de congresistas, como al usar un congress_number bajo (por ejemplo, 1), optamos por no crear un miembro ficticio para que el archivo XML valide con el XSD. En lugar de esto, mostramos un mensaje en el HTML indicando que no se encontró información para esos congresistas.

A lo largo del desarrollo del proyecto, nos enfrentamos a varias dificultades. En primer lugar, fue necesario familiarizarnos nuevamente con la sintaxis de Bash, ya que hacía tiempo que no la utilizamos, y esto nos tomó tiempo adicional. Además, al generar el archivo XML con XQuery, enfrentamos problemas de duplicación de etiquetas, lo que requirió ajustes para obtener el formato correcto. Otra dificultad técnica fue decidir cómo representar periodos múltiples para congresistas con varios términos en una cámara (House o Senate), y finalmente elegimos mostrar el from del primer startDate y el to del último endDate para simplificar la visualización. También nos encontramos con la falta de imágenes para algunos congresistas en la API, lo que dejaba espacios vacíos en la tabla HTML; para mantener la uniformidad, decidimos mostrar una "X" en lugar de la imagen, conservando el tamaño 50x50 píxeles. Finalmente, en XSLT, enfrentamos complicaciones con los periodos de personas cuyo mandato no había terminado; para estos casos, omitimos el To y mostramos solo el From en la tabla HTML, evitando así espacios vacíos.

Además de las tareas previamente planeadas, tuvimos que investigar y aprender conceptos que no se habían abordado en clase. En primer lugar, aprendimos a realizar consultas a una API pública, algo que no habíamos practicado en otras materias, lo cual nos permitió trabajar con datos reales. Por otra parte, esta fue nuestra primera experiencia en la creación de una página web para presentar datos, lo cual nos introdujo al formato web y al diseño básico de interfaces. También fue necesario investigar cómo implementar un manejo de errores en Bash y XQuery para asegurar que nuestro flujo de trabajo respondiera adecuadamente a diferentes condiciones.

Cada integrante asumió roles específicos para lograr una división eficiente del trabajo. Eduardo Tormakh y Agustín Korman fueron los responsables de la consulta XQuery, encargándose de la creación y depuración de "extract_congress_data.xq" para generar el archivo "congress_data.xml". Rodrigo Alejandro Hernández asumió el rol de responsable de la plantilla XSLT, desarrollando "generate_html.xsl" y asegurando la correcta presentación de los datos en formato HTML. Además, Rodrigo Alejandro Hernández se desempeñó como responsable del funcionamiento global del proyecto, supervisando la integración de todas las partes y garantizando la coherencia entre los componentes del flujo de trabajo. Por último, Eduardo Tormakh fue responsable de la presentación, ocupándose de la redacción y organización del informe, documentando cada paso del proyecto.

CONCLUSIÓN

Para concluir, este trabajo práctico nos permitió aprender a extraer datos de una API pública, algo que no habíamos hecho en otras materias y que nos dio la oportunidad de trabajar

con información real. Fue interesante ver cómo datos en bruto se podían transformar y organizar para luego ser utilizados en algo concreto, como una página HTML.

Además, esta fue la primera vez que trabajamos con HTML, lo cual nos permitió acercarnos al mundo web, un área que no habíamos explorado hasta ahora. Nos dio una idea básica de cómo presentar información en una página de forma accesible y estructurada.

Un desafío importante fue coordinar el trabajo en equipo, ya que varias partes del proyecto dependían entre sí. Organizar nuestras tareas y conectar los componentes para que funcionaran juntos requirió esfuerzo y comunicación, pero logramos cumplirlo. Esta experiencia nos ayudó a entender mejor cómo conectar varios lenguajes y poder presentarlos de manera efectiva para un potencial usuario y pudimos entender mejor la conexión de distintas tecnologías y la manipulación de datos en formato XML.