



Instituto Tecnológico de Buenos Aires

Diseño y Procesamiento de
Documentos XML
Trabajo Práctico Especial

Integrantes Grupo 8:

Gonzalo Martin Elewaut 61212

Lucas Agustín Ferreiro 61595

Francisco José Domínguez Machado 61068

1. Introducción

El NASCAR (o “National Association for Stock Car Auto Racing”) es la categoría automovilística más popular de Estados Unidos; y también es considerada la competición de automóviles de serie más importante del mundo. En este trabajo se busca extraer y presentar datos e información acerca de los distintos conductores que participaron en cierto año y en cierta competencia. Para ello se utiliza la API oficial de NASCAR brindada por SportRadar (https://developer.sportradar.com/docs/read/racing/NASCAR_v3); y el uso de distintos lenguajes para la escritura, validación y transformación de los documentos a utilizar (entre los cuales se destacan XML, XSLT, XQuery y XML Schema) y un script bash *tpe.sh* para la ejecución automatizada de las instrucciones, así como también la validación de distintos parámetros.

2. Desarrollo

El trabajo consiste en un archivo de extensión .sh (script bash) encargado de la automatización de todo el proceso. Dicho script debe ser invocado en la entrada estándar de una consola de comandos con dos parámetros (año y categoría), como se muestra a continuación *tpe.sh (año) (tipo de categoría)*. Un ejemplo de invocación: *tpe.sh 2019 mc*.

El ejecutable es el encargado de validar los parámetros ingresados, y en el caso de fallar en alguna de las validaciones se genera un documento *nascar_data.xml*, donde el nodo raíz es `<nascar_data>` y este mismo posee un único nodo hijo `<error>`, el cual tendrá un texto adecuado acorde al tipo de error. A continuación, se listan las validaciones y el texto del nodo `<error>`:

- ✓ Que la cantidad de parámetros sean exactamente dos, siendo primero el año y luego el tipo de competencia. ***“There must be 2 arguments (year, serie_type)”***
- ✓ Que el año ingresado sea un número entero positivo y que se encuentre entre 2013 y 2021 inclusive. ***“First parameter ‘Year’ must be an integer between 2013 and 2021”***
- ✓ Que la categoría sea únicamente alguna de las siguientes opciones: ‘sc’, ‘xf’, ‘cw’, ‘go’, ‘mc’ o ‘enas’. ***“Second parameter ‘Serie_Type’ must be only one of the following: ‘sc’, ‘xf’, ‘cw’, ‘go’, ‘mc’ or ‘enas’”***
- ✓ También existe un caso particular, que es el siguiente: el tipo de competencia ‘enas’ valida si y sólo si el año es 2020. ***“Serie_Type ‘enas’ (eNASCAR) is only supported with 2020 season”***

En cualquier caso, se verifica que el documento valide con *nascar_data.xsd*, para luego generar la página Markdown *nascar_page.md* donde solamente figurará el error cometido al ejecutar el script *tpe.sh*.

Una vez realizadas todas las validaciones y comprobar que los parámetros sean correctos, se proceden a descargar los datos de los corredores y sus estadísticas por medio del comando 'curl' (en el cual se utiliza el año y el tipo de serie que fueron ingresadas, además de una variable de entorno SPORTRADAR_API, seteada previamente con la key obtenida de la página SportRadar), y el flag -o, para obtener dichos datos en los archivos *drivers_list.xml*, *drivers_standings.xml*. Inmediatamente se utiliza el comando 'sed' para remover los XML Schema NameSpace de los documentos descargados. Luego el script utiliza un archivo XQuery llamado *extract_nascar_data.xq* para filtrar los datos de los campos relevantes, así como también ordenar alfabéticamente por nombre a los corredores; dichos datos se almacenan en *nascar_data.xml* y su formato es validado por el archivo *nascar_data.xsd* (brindado por la cátedra). Finalmente, el script se encarga de transformar los datos obtenidos en *nascar_data.xml* a una página Markdown (*nascar_page.md*) mediante el uso del archivo *generate_markdown.xsl*.

2.1 Dificultades encontradas

Encontramos dificultades para machear los atributos id de *drivers_list.xml* con los de *drivers_standings.xml*, con el objetivo de conseguir las estadísticas y el rango de cada conductor. Pero finalmente, utilizando funciones y pensándolo entre todos los integrantes del trabajo, se pudo resolver sin problemas. Finalmente, concluimos que, si un conductor no poseía un rango, los nodos de statistics se encontrarían vacíos.

Además, al momento de validar el documento *nascar_data.xml* con el archivo XQuery *nascar_data.xsd*, este mismo nos arrojaba un error llamado *java.net.MalformedURLException: no protocol*.

Por último, nos encontramos con la dificultad para llevar a cabo el script bash *tpe.sh*, pues no recordábamos la sintaxis requerida, ni tampoco como lograr que todo el proceso sea automático.

2.2 Aspectos que requirieron de investigación

Aquellos temas que requirieron de investigación fueron:

- El uso del comando 'sed' para poder eliminar los XML Schema NameSpace de los documentos *drivers_standings.xml* y *drivers_list.xml* y no tener inconvenientes en el desarrollo del trabajo.
- Fue necesario buscar información acerca de la creación de un script bash para poder validar los parámetros y manejar los errores mencionados anteriormente. Además de

realizar los comandos para generar los documentos XML tales como *drivers_standings.xml*, *drivers_list.xml*, *nascar_data.xml* y el de la página Markdown *nascar_page.md*.

- Se requirió investigar información sobre el error *java.net.MalformedURLException: no protocol*, para ver cuál era su significado.
- Por último, investigamos el diseño y formato de una página Markdown, para conseguir que esta misma corresponda con lo solicitado.

2.3 Roles de cada integrante

Los roles de cada integrante fueron los siguientes:

- Responsable del funcionamiento de la consulta XQuery: Francisco Domínguez
- Responsable del funcionamiento de la plantilla XSLT: Gonzalo Elewaut
- Responsable del funcionamiento global del proyecto: Lucas Ferreiro
- Responsable de la presentación: Lucas Ferreiro

3. Conclusión

Para concluir este trabajo, se podría mencionar que logramos aplicar los conocimientos adquiridos en el curso y afianzamos aquellos aspectos que no logramos comprender a lo largo del cuatrimestre. Nos complementamos como grupo en la búsqueda de las dificultades encontradas y resolvimos dichos aspectos, cumpliendo así los objetivos que nos propuso este trabajo.