



Trabajo Práctico Especial

“Diseño y Procesamiento de Documentos XML”

Agustina Sanguinetti (63115) - msanguinetti@itba.edu.ar

Sol Rodriguez (63029) - solrodriguez@itba.edu.ar

Micaela Perillo (62625) - miperillo@itba.edu.ar

Uriel Arias (63504) - uarias@itba.edu.ar

Grupo 12

15 de Noviembre, 2022

Instituto Tecnológico de Buenos Aires

1. Introducción

El objetivo de este trabajo es aplicar las herramientas de consulta y transformación de documentos XML con el fin de obtener una página Markdown con el nombre y fechas de una determinada temporada, junto con sus fases, grupos y lista de competidores y partidos. Estos datos se obtienen a partir de SportRadar, un sitio web que permite consumir datos de diversos deportes en forma de invocaciones a API REST.

2. Desarrollo

2.1. Trabajo Realizado

En primer lugar, utilizando la API de SportRadar para Fútbol, generamos los archivos `season_info.xml` y `season_summaries.xml`. A partir de los datos provistos por estos dos archivos implementamos la consulta XQuery con la estructura indicada en el XML Schema `season_data.xsd`. La salida de esta consulta es un archivo XML `season_data.xml`, cuyo contenido se compone únicamente de la información de nuestro interés. A partir de ello, desarrollamos un XSLT `generate_markdown.xsl`, para poder luego generar el archivo `season.md`, el cual permite visualizar la información deseada. Por último, para que todo este proceso se ejecute automáticamente, creamos el script de bash `tpe.sh` que genera la página Markdown de salida a partir de las transformaciones realizadas. Este mismo recibe como parámetros el id de la temporada, la cual es validado mediante el uso de errores.

2.2. Dificultades

Una de las mayores dificultades surgió a la hora de generar los archivos XML a partir de los llamados a la API. La primera vez que pedimos la API Key, luego de varias horas de espera, ésta nunca se activó. Es por ello que optamos por solicitar otra desde otra cuenta pero también tuvimos inconvenientes. Por fortuna, era un problema de nuestro código, así que lo pudimos resolver.

Quizás el mayor desafío fue escribir el archivo `extract_season_data.xq`. En primera instancia, no fuimos capaces de lograr que se ejecuten los nodos “event” y sus nodos descendientes. Al principio creímos que se trataba de un problema generado por un error de escritura. Sin embargo, llegamos a la conclusión que el problema surgía al querer ingresar al archivo `seasons_summaries.xml`. Fue entonces cuando nos dimos cuenta que el comando “sed” no estaba alterando al archivo y, por lo tanto, no estaba quitando el Namespace. Una de las soluciones que encontramos fue ejecutar “sed -i”, mas esto nos generaba el error:

“unterminated substitute pattern”. Por lo tanto, optamos por crear el archivo `remove_namespace.sh` y logramos solucionar el problema.

Cabe destacar que este inconveniente no solo retrasó la elaboración del archivo `extract_season_data.xq`, sino que también dificultó comprobar el buen funcionamiento del archivo `tpe.sh` y `generate_markdown`.

2.3. Investigación

Para la última parte del trabajo, mostrar los datos en la página Markdown, se investigó acerca de cómo pasar los datos procesados a una tabla.

Por otra parte, fue necesario buscar ciertos comandos de bash, como también investigar acerca del funcionamiento del comando `sed` que, como ya mencionamos, nos nos estaba generando el resultado esperado.

Para la elaboración del archivo de xQuery, investigamos acerca de “`data()`”.

2.4. Roles

En cuanto a los roles, si bien el trabajo fue en equipo, hubo una división de las tareas. Agustina y Sol se encargaron de la realización de la consulta XQuery, la cual produce como salida el documento XML que contiene los datos deseados. Por otro lado, Micaela se ocupó de la creación de la transformación XSLT para producir la página Markdown, y por último Uriel se encargó del script bash `tpe.sh` para que todo el proceso se ejecute automáticamente. La escritura del informe fue realizada en conjunto, así como las dificultades que se presentaron en el desarrollo del trabajo fueron evaluadas con todos los integrantes.

3. Conclusiones

Con los conocimientos adquiridos en la materia, como XML y XSD, y el uso de Bash es posible procesar grandes cantidades de datos con el fin de mostrarlos de manera más amigable en otro formato, como por ejemplo HTML, PDF, Markdown, Tex, entre otros. En el caso del trabajo presentado, procesamos datos obtenidos a través de la API de SportRadar, específicamente de Fútbol y los mostramos en una página hecha en Markdown para que puedan ser vistos de forma clara.