

Lab-Book

Bases de Datos

Documentales (XML)

Beginning 21-03-2022

Jesús Baltasar Fernández

Carlos Gómez Fernández

Domenico Fasciano

Bases de Datos Avanzadas

Contents

Lunes, 21 Marzo 2022	5
1 Introducción	5
1.1 Hito 1	5
1.2 Hito 2	5
1.3 Hito 7	6
Martes, 22 Marzo 2022	7
1 Introducción	7
1.1 Hito 3	7
Jueves, 24 Marzo 2022	9
1 Introducción	9
1.1 Hito 3	9
Lunes, 28 Marzo 2022	11
1 Introducción	11
1.1 Hito 4	11
1.2 Hito 5	11
Martes, 29 Marzo 2022	13
1 Introducción	13
1.1 Hito 4	13
1.2 Hito 5	13
1.3 Hito 6	13
Miércoles, 29 Marzo 2022	15
1 Introducción	15
1.1 Hito 8	15
Dificultades	17

Lunes, 21 Marzo 2022

1 Introducción

Comenzará el desarrollo del trabajo sobre Bases de Datos Documentales (XML), consensuando el *Hito 1*, correspondiente con la selección de tecnologías a emplear, el *Hito 2* correspondiente con la selección de las categorías y temas de los trabajos a realizar e iniciando el *Hito 7*, correspondiente con la elaboración del cuaderno de laboratorio.

1.1 Hito 1

Tras discutir las tecnologías necesarias para el desarrollo del trabajo, se ha optado por emplear inicialmente Entornos de Desarrollo Integrado como *Microsoft Visual Studio 2019*, *Microsoft Visual Studio Code* y *Eclipse*, lenguajes de programación como *Java*, *Python* y *XSLT*, y como motor de Base de Datos Documental *eXist-db*.

1.2 Hito 2

En este hito, seleccionaremos tanto las categorías de las preguntas recogidas en el archivo aportado por el profesor, las cuales serán empleadas para la realización de la primera parte del proyecto, como el tema del trabajo para la realización de la segunda parte del proyecto. En cuanto a las categorías, para la primera parte del proyecto, de las preguntas seleccionadas son las siguientes: *Business & Finance*, *Cars & Transportation*, *Computers & Internet* y *Entertainment & Music*. En cuanto al tema, para la segunda parte del proyecto, se seleccionó el correspondiente con los fondos de inversión, obteniendo la información necesaria para su desarrollo de la *web* sobre la *Comisión Nacional del Mercado de Valores (CNMV)*.

1.3 Hito 7

Comenzará la elaboración del cuaderno de laboratorio, en el cual se plasmarán los hitos más importantes durante el desarrollo del trabajo, haciendo uso de la plantilla *Latex* proporcionada por el profesor.

Martes, 22 Marzo 2022

1 Introducción

Iniciando el *Hito 3*, correspondiente con la generación del archivo con el esquema *XSD* a partir del archivo *XML* proporcionado.

1.1 Hito 3

El inmenso tamaño del archivo *XML* hace imposible el trabajar directamente con él, por lo que se optó por generar un programa en *Java* haciendo uso del *IDE Eclipse* para extraer de él las preguntas cuya categoría se correspondía con las seleccionadas en el *Hito 2*.

Jueves, 24 Marzo 2022

1 Introducción

Finalizando el *Hito 3*, correspondiente con la generación del archivo con el esquema *XSD* a partir del archivo *XML* proporcionado.

1.1 Hito 3

Tras generar un nuevo archivo *XML* únicamente con las preguntas de las categorías seleccionadas, nos resultó más sencillo la generación del archivo *XSD* con el esquema *XML* que estructura el archivo proporcionado. Este fue generado haciendo uso de una *web* incluida en la bibliografía de este mismo documento (*XML Schema Definition*).

Lunes, 28 Marzo 2022

1 Introducción

Iniciando el *Hito 4* y el *Hito 5*, correspondientes ambos con el uso del motor de bases de datos documental.

1.1 Hito 4

Comenzamos con la instalación del motor de bases de datos documental y *noSQL*, seleccionado en el *Hito 1*, el cual es *eXist-db*, un proyecto de código libre, el cual emplea *Java* como lenguaje base de programación. Aunque existe bastante documentación del mismo nos resultó imposible realizar su instalación en *Windows*, aun teniendo *Java* instalado en nuestras máquinas y ser versiones *Windows* diferentes (*Windows 10* y *Windows 11*).

1.2 Hito 5

La imposibilidad de instalación del motor de base de datos documental *eXist-db* nos imposibilitó el inicio del *Hito 5*, correspondiente con la carga de los archivos *XML* filtrados por categorías en el mismo, hecho que ha supuesto un retraso en la planificación del proyecto.

Martes, 29 Marzo 2022

1 Introducción

Finalizando el *Hito 4* y el *Hito 5*, correspondientes ambos con el uso del motor de bases de datos documental y realización del *Hito 6*, correspondiente con las consultas sobre los datos almacenados y la obtención de conclusiones.

1.1 Hito 4

Tras encontrarnos en la tesitura de no poder completar el proyecto exitosamente, decidimos recurrir al uso de *Virtual Machine*, dándonos la posibilidad de realizar la instalación del motor de base de datos documental en otro sistema operativo. Seleccionamos *Linux*, ya que, estamos familiarizados con él, concretamente en *Ubuntu*, tras realizar la correspondiente creación de la *Virtual Machine*, iniciamos la instalación de *eXist-db* de la misma forma que hicimos en *Windows* y con los mismos requisitos, pero esta vez de manera satisfactoria.

1.2 Hito 5

Tras la instalación y correcto funcionamiento del motor de bases de datos documentales, se comenzó con la carga de los archivos *XML* generados al filtrar el original por las categorías seleccionadas.

1.3 Hito 6

Finalmente, sobre estos archivos se realizaron una serie de consultas haciendo uso del lenguaje de consultas, acorde con el tipo de archivos cargados en el motor de base de datos documental, en este caso, al tratarse de archivos *XML*, se optó por emplear el lenguaje de consultas *XQuery*, y a partir de los resultados de dichas consultas, obtener conclusiones tanto de los datos, como del proyecto desarrollado.

Miércoles, 29 Marzo 2022

1 Introducción

Desarrollo del *Hito 8*, correspondiente con la elaboración de la presentación.

1.1 Hito 8

Realización de la presentación sobre el trabajo desarrollado, así como del análisis de las ventajas que proporcionana las bases de datos documentales que almacenan archivos *XML*, haciendo uso de la herramienta *PowerPoint*.

Dificultades

Entre las principales dificultades encontradas durante el desarrollo del trabajo ha sido el uso del motor de base de datos documentales *eXist-db*, ya que, nos fue imposible su instalación en *Windows*, teniendo que hacer uso de una *Virtual Machine* con sistema operativo *Linux*, con interfaz gráfica (*Ubuntu*).

Bibliography

- [1] Instituto de Tecnología de Massachusetts (MIT), *XML Schema Definition* - <http://www.xml-tools.net/schemaviewer.html>. Massachusetts, Estados Unidos, 2018.
- [2] Pharos, *XML: Leer y escribir archivos XML en Python*. Estados Unidos.
- [3] Phyton, *XML: La API XML de ElementTree*.
- [4] TMaxEcLearning, Fernando Pullutasig *XML: EXTRAER DATOS DE FACTURA XML (formato del SRI EC) USANDO PYTHON*. 3 de Marzo del 2021.
- [5] w3schools, *XML Tutorial*.
- [6] eXist-db, *XML: http://exist-db.org/exist/apps/homepage/index.html*. 2018.