

5. Metodología

En este capítulo⁴ se describen los pasos que se siguieron para desarrollar el constructor automático de la base de datos, y la posterior población de dicha base. Se describen todas las especificaciones, requisitos y modelado. Obteniéndose de este análisis los diferentes diagramas UML.

Descripción de lo que se requiere

El objetivo de esta tesis era desarrollar un constructor automático de bases de datos relacionales a partir de esquemas XML para ser pobladas con datos contenidos en los documentos XML. Los pasos que se siguieron para elaborar tal proyecto, se basaron en la metodología de la ingeniería de software, tomando en cuenta las necesidades que se requerían y cuales eran los elementos con que se contaba para tomarlos como base para desarrollar dicho constructor. Se abordaron diferentes aspectos que se tenían que considerar. Para facilitar este análisis se enlistaron en los requisitos las características mínimas que debía tener el programa, y se establecieron como prerrequisitos a los documentos o el sistema con los que se debería trabajar. Enseguida se muestran las listas de los requisitos y prerrequisitos, usados para la construcción y población de la base de datos.

⁴ La información presentada en este capítulo se obtuvo de revisar las siguientes fuentes: Alarcón, Raúl, *Diseño Orientado a objetos con UML*, Eidos 2000; Fowler, Martin, *UML gota a gota*, Addison Wesley 1999; Pressman Roger, *Ingeniería del Software*, McGraw-Hill 2002 y Zapata J. Carlos Mario, *Ingeniería del Software: Una disciplina de modelamiento*, Colombia, 2006. Además se consultaron las páginas siguientes:

- <http://es.tldp.org/Tutoriales/doc-modelado-sistemas-UML/doc-modelado-sistemas-uml.pdf>
- <http://www.ingenierosoftware.com/analisisydiseno/uml.php>

Prerrequisitos

Los prerrequisitos se refieren principalmente a la forma en que se encuentran elaborados los documentos XML. Para construir la base de datos, la idea es tomar la información de su estructura codificada en el esquema XML que se utiliza para determinar si los documentos están bien formados y validados.

Requisitos

Los requisitos se refieren a los resultados que se quieren obtener, tanto para el constructor como para el alimentador, y están resumidos en los siguientes puntos:

- Construir automáticamente una base de datos bibliográfica para el Corpus Histórico del Español en México que está constituido por un conjunto de documentos XML con información variada, además de la bibliográfica.
- A partir de un esquema XML y documentos XML, validados por dicho esquema, obtener la información necesaria para crear una base de datos y posteriormente poblarla.
- No es necesario que se cree la base de datos a partir del nodo inicial del esquema, sino que puede empezar a crearse dicha base a partir de un nodo secundario.
- La base de datos creada será SQL (PostgreSQL)

Desarrollo

Una vez que se definieron cómo eran los tipos de documentos a los cuales se les iba a extraer la información (los prerrequisitos) y qué es lo que debería obtenerse como resultado de dicha extracción de información (los requisitos), se procedió a plantear las mejores alternativas para poder llegar a un resultado óptimo.

Para lograr un resultado adecuado, como metodología utilicé un modelado estructurado, ayudándome de los pasos de desarrollo de software y más específicamente basándome en el lenguaje UML; que aunque no es una metodología formal propiamente, sí proporciona diferentes diagramas que facilitan el diseño.

Para tener una idea de la complejidad del programa a desarrollar, lo primero que hice fue realizar los diagramas de Casos de Uso usando la notación de UML. El Diagrama de Casos de Uso sirve para analizar de una manera general las exigencias que se debe cumplir para realizar el constructor de la base de datos relacional, ya que este tipo de diagramas tiene dos componentes básicos que ayudan mucho en el diseño: los actores que son la representación de las funciones que pueden realizar los diferentes usuarios del sistema y los casos de uso que simbolizan una función específica que realiza el sistema para solucionar una acción específica. Estos ayudan a reconocer cómo son las interacciones entre los diferentes usuarios o actores y el sistema; todo esto a partir de los análisis de requisitos y prerrequisitos (arriba mencionados).

Hay que mencionar que además de los diagramas de Casos de Uso, también se utilizaron otros diagramas de la metodología UML; que se mencionan a continuación:

- Diagrama de secuencia
- Diagrama de colaboración
- Diagrama de clases
- Diagrama de estados

Estos diagramas se encuentran en el apéndice: Diagramas UML.

Una vez que se logró identificar por medio de los diagramas UML las diferentes rutinas que debería de tener el constructor y el alimentador de la base de datos, el siguiente paso fue analizar las características tanto de los documentos XML y esquemas XML, obteniendo las siguientes conclusiones:

Los documentos XML tienen la información clasificada en piezas de datos delimitadas por etiquetas, aunque existen etiquetas “nulas”(<!--comentario -->), que pueden contener comentarios para facilitarnos como usuarios de los documentos y esquemas XML la identificación de ciertas etiquetas, sin embargo dichas etiquetas “nulas” no son consideradas parte de un documento XML y por lo tanto no representan información útil para la creación y alimentación de la base de datos. También por facilidad en los documentos y esquemas XML se utilizan espacios y tabulaciones para darle un orden a un esquema o documento XML, pero tales elementos (espacios y tabulaciones) no son información pertinente a este trabajo, por lo que hay que ignorarla.

Se asume que los documentos XML que se van a utilizar para alimentar la base de datos ya están bien formados y validados. Como se mencionó en el capítulo de *Metodología XML*, un esquema XML consiste en las reglas que deben seguir los documentos XML, dichas reglas de un esquema (archivo tipo XSD) nos indican el orden de los elementos, los valores de los datos de atributos y elementos, o la multiplicidad de valores de un documento (por ejemplo elementos que deben repetirse o que no pueden repetirse), etc.

Los documentos XML tienen una estructura jerárquica, en la que existe sólo un único elemento raíz en el que todos los demás están contenidos; y todas las etiquetas están correctamente anidadas una dentro de otra.

Después de analizar y observar las características de los documentos y esquemas XML, el siguiente paso fue analizar cuál era la mejor forma de leer los documentos XML, conocer cuáles eran las posibles opciones para recorrerlos y finalmente el resultado que se podría obtener al realizar el recorrido del esquema para crear la base de datos.

El resultado del análisis que se realizó en este capítulo, tanto de los diagramas UML, como de las características más importantes de los documentos XML, dieron como resultado los diferentes pasos que debería seguir para la elaboración del constructor y alimentador automático de la base de datos.

En el siguiente capítulo se muestra a detalle la forma en que se diseñó el constructor automático de la base de datos, tomando diferentes características a considerar.