

8. Conclusiones finales

En el presente trabajo, lo que hice para elaborar el constructor fue primero investigar lo que eran los esquemas XML; luego, visualizar la posible solución de un caso en particular y; finalmente, decidir cuál era el comportamiento que tenía que seguir el constructor cuando encontrara las diferentes etiquetas dentro del esquema, lo que forma una idea aproximada de cómo debería ser la base de datos que se quería construir.

Después de saber cuáles eran las posibles opciones que se podían seguir, se buscó saber cuáles eran las diferentes formas en que se podría presentar la información dentro del esquema XML; como por ejemplo, una etiqueta podría empezar en un renglón y terminar en el siguiente. El constructor tendría que reconocer cuáles eran esos casos y juntar la etiqueta en un solo renglón. También era importante eliminar tabulaciones, entre otras marcas de espacio en blanco, y sustituirlas por espacios en blanco.

Sabiendo todo lo anterior, hice un análisis y obtuve una idea general de cómo debería estar estructurado el programa, y decidí qué clases se necesitaban crear para poder realizar el constructor de la base de datos, ayudado de los diagramas UML y el modelado estructurado.

Lo que siguió fue, tomando como referencia el recorrido que se hizo de un esquema XML (que se mencionó al principio), hacer el análisis más general y saber cuáles eran las posibles decisiones que podría tomar el constructor, tanto en

la lectura de los datos del mismo archivo de texto, como en las decisiones que debe tomar para crear las tablas. Fue también importante encontrar una forma para que el proceso sea recursivo y avanzar en la estructura del esquema para crear las tablas y subtablas para, finalmente, crear los comandos SQL, ejecutarlos y crear la base de datos.

Las desventajas de todo esto son que no se abarcan todos los posibles casos que puede haber al momento de crear un esquema XML y un documento XML, debido a que el desarrollo es específico para el Corpus Histórico de Español en México. La otra desventaja importante es que se generan algunas tablas vacías. De hecho, con estas tablas vacías la normalización de la base de datos se vuelve muy difícil.

Lo importante es que con este desarrollo se pueden crear bases de datos, generando simples reglas a partir de un esquema XML. Lo cual ayuda especialmente en caso de no tener muchos conocimientos acerca de este tema, y se cuente ya con una gran cantidad de documentos XML, y también que las instrucciones generadas con el constructor se puedan ejecutar en cualquier gestor de base de datos PostgreSQL, o SQL.

El desarrollo del trabajo descrito en esta tesis soluciona un problema específico de los corpus electrónicos del Instituto de Ingeniería, específicamente del Corpus Histórico del Español en México. Como trabajo futuro y ya que estos corpus están constituidos por documentos XML, valdrá la pena explorar la

consulta de los datos de estos documentos, especialmente los bibliográficos, mediante las estrategias desarrolladas específicamente para este tipo de documentos; esto es, la aplicación de las tecnologías relacionadas con xQuery que no requieren la construcción adicional de una base de datos. Cabe mencionar que incluso PostgreSQL ya cuenta con comandos específicos para consultas en documentos XML.