

INSTITUTO TECNOLÓGICO DE COSTA RICA

TAREA #1

Saúl Zamora

profesor
Kevin Moraga

March 1, 2017

1 Introducción

El objetivo principal de esta tarea es la introducción al análisis de bases de datos ya existentes. Para lo cual se lleva a cabo un desarrollo dividido en dos partes.

1.1 Parte 1

En la primera parte, se desarrolla un modelo de base de datos relacional para una base de datos encargada de organizar duelos del conocido juego de cartas *Magic The Gathering*.

1.2 Parte 2

La segunda parte consiste en el análisis de una base de datos existente. Utilizando la base de datos pública proveída por el profesor, se debe realizar un modelo de base de datos entidad-relación de la misma, junto con una serie de consultas SQL.

Además se pide la implementación de un cliente en consola escrito con el lenguaje Perl para GNU/Linux con las siguientes funcionalidades:

- Realizar consultas a <https://www.wikileaks.org/hackingteam/emails/>
- Parsear los resultados de las consultas a un documento CSV.
- Permitir realizar consultas por:
 - Palabra en el correo.
 - Correo de envío.
 - Correo de recibido.
- Permitir realizar búsquedas dada una expresión regular utilizando los tipos de consulta anteriores.
- Utilizar la biblioteca *GetOps* para estilizar el comando de búsqueda en consola.

2 Ambiente de desarrollo

Para la elaboración de los diagramas de entidad-relación de bases de datos solicitados se utilizó la siguiente herramienta:

- StarUML v2.8.0

Para la elaboración de consultas SQL se hizo uso de la siguiente herramienta:

- MySQL community for Windows v5.7.17

Todo ejecutado sobre un sistema operativo Windows 10 Home Edition x64.

Para la elaboración del cliente de consultas de Perl, se utilizó el siguiente ambiente:

- Perl v5.22.1 en el SO GNU/Linux Ubuntu 16.04 LTS

3 Estructuras de datos usadas y funciones

Para el cliente en Perl se crearon dos rutinas principales:

- *printOptions()*: retorna la ayuda del programa (Acerca de...).
- *makeRequest()*: realiza la consulta deseada a WikiLeaks y además parsea los resultados a un archivo “results.csv”.

Se utilizan las siguientes librerías:

- *Getopt::Std* : obtiene y parsea las opciones de consola para realizar las consultas adecuadamente.
- *CGI::Carp* : captura de errores.
- *LWP::UserAgent* : para realizar consultas en navegadores.
- *HTTP::Request* : para realizar las consultas en sí.
- *HTML::TableExtract* : extraer y parsear tablas HTML con los resultados de las consultas.

4 Instrucciones de ejecución'

Para la ejecución del cliente en Perl de la segunda parte, es necesario ejecutar el siguiente comando para instalar la librería *HTML::TableExtract*:

- *sudo apt-get install libhtml-treebuilder-xpath-perl* : paquete necesario para la extracción de tablas en HTML.

Para ejecutar el cliente se debe seguir los siguientes pasos:

- Abrir una ventana de consola.
- Navegar hasta el folder que contiene el archivo “cliente.pl”.
- Ejecutar el comando “perl cliente.pl -t <tipo-consulta> <texto>” donde:
 - *<tipo-consulta>* : puede ser “email_from”, “email_to” o “email_content”.
 - *<texto>* : si la primera opción es “email_from” o “email_to”, esta debe ser un email válido o un dominio de email válido. Si la primera opción es “email_content”, esta será el texto a buscar en el cuerpo del email.

5 Bitácora de trabajo

- 25-02-2017:
 - 1 hora - modelado de base de datos de Magic (texto).
 - 1.5 horas - modelado de base de datos de Magic (StarUML).
 - 1 hora - instalación y configuración de MySQL en Windows.
 - 3 horas - investigación sobre cliente en Perl. Consultas a Wikileaks.
 - 1 hora - investigación de GetOps en Perl.
 - 1 hora - documentación.
- 27-02-2017:
 - 1 hora - primer intento al cliente en Perl.
 - 2 horas - Extracción de tablas del HTML en Perl.
- 28-02-2017:
 - 2 horas - Remover cuerpo del email para evitar la salida desordenada. Agregar archivo de salida CSV.
- 01-03-2017:
 - 4 horas - Agregar opciones a la aplicación de consola. Configurar los tipos de consulta en el cliente.
 - 1 hora - documentación.

Total de horas trabajadas: 18.5 horas.

6 Comentarios finales

- Debido al horario laboral, la falta de tiempo fue una limitante y no fue posible realizar las consultas sobre la base de datos MySQL de la tarea.
- El cliente de Perl realiza exitosamente las consultas por “email_from” e “email_to”; la consulta por texto cualquiera está correctamente estructurada, según lo demostraron pruebas en browser, pero aún así no retorna resultados.

7 Conclusiones

- Perl es un lenguaje muy útil a la hora de manipular grandes cantidades de información de texto, como lo es el HTML.
- Analizar el trabajo realizado por otros, como la base de datos MySQL de esta tarea, es un trabajo tedioso y lento el cual requiere tiempo.

References

- [1] Cmo jugar magic the gathering. Retrieved February 25, 2017, from <http://es.wikihow.com/jugar-Magic-The-Gathering>
- [2] Getopt:Long. Retrieved February 26, 2017, from <http://perldoc.perl.org/Getopt/Long.html>