

Escuela Superior Politécnica del Litoral

Facultad de Ingeniería Electricidad y Computación

Lenguajes de Programación

Parseo de Datos Personales

Proyecto Parcial

I Término 2014-2015

Realizado por: Raquel Villón Ramírez & Marcelo Sánchez Vega

Introducción

Perl es un lenguaje de programación interpretado, es decir que ejecuta el código línea por línea sin previa compilación.

Este lenguaje fue creado principalmente para el procesamiento de texto y la elaboración reportes, de ahí su nombre, por sus siglas en inglés Practical Extraction and Report Language (Perl). Este lenguaje hace uso de las expresiones regulares, para realizar búsquedas dentro de grandes sets de información.

Descripción

Este proyecto es un programa realizado en el lenguaje de programación Perl, el cual se encarga de leer un archivo de texto plano (*.txt) y por medio de expresiones regulares encuentra varios datos personales de múltiples individuos.

Objetivos Alcanzados

En este proyecto nosotros logramos obtener la siguiente información de un dataset de información.

Datos encontrados:

- Nombre
- Edad
- Estatura
- Peso
- Teléfono
- Correo
- Carrera
- Universidad
- Estado Civil
- Lugar de Nacimiento

Objetivos Inconclusos

Aunque se lograron alcanzar los objetivos acordados en clase, la forma específica en que la información es obtenida, está parcialmente ligada al dataset hecho en clases, puesto que cada expresión regular fue elaborada para encontrar todas las coincidencias del dataset propuesto para el presente proyecto.

No se ha podido obtener datos acordados como opcionales en el proyecto.

Breve descripción del programa

Este proyecto fue desarrollado haciendo uso de PADRE, un editor de texto que da la facilidad de ejecutar código en perl mediante el uso de la ventana de comandos (cmd).

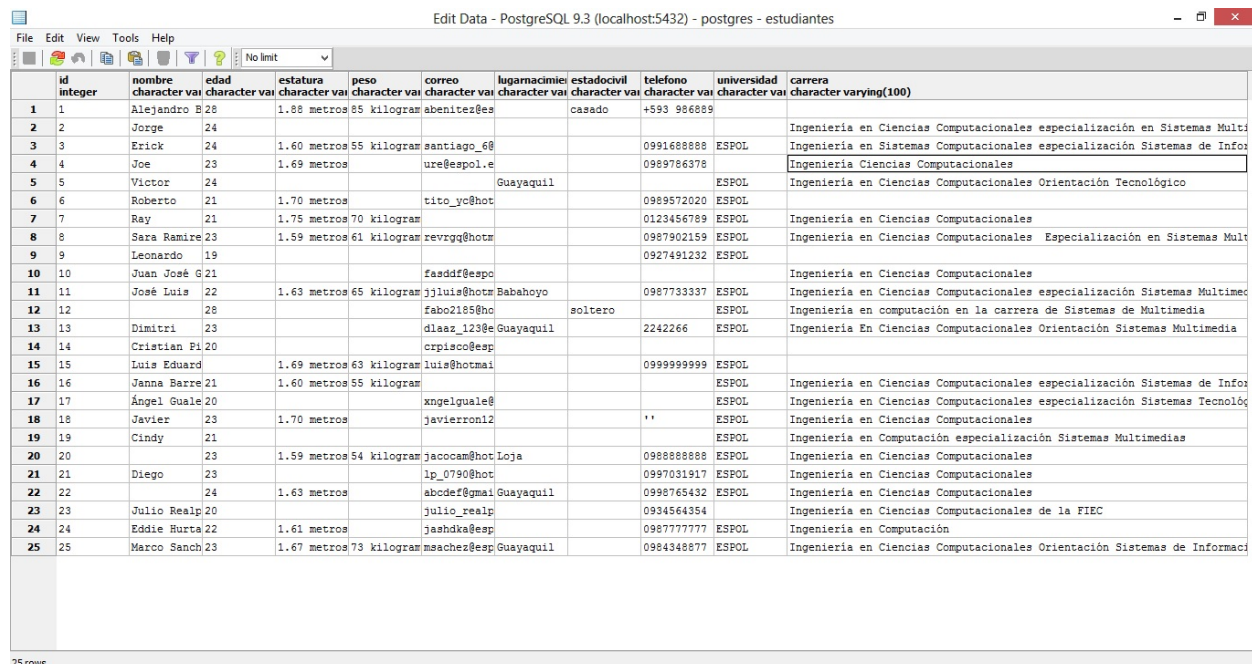
El programa puede ser ejecutado llamando al archivo “*main.p*”, el cual llamará inmediatamente al archivo “*dataset.txt*” que es el que contiene el set de información para procesar.

Luego de abrirlo y guardar su contenido dentro de un arreglo, se procede a analizar su contenido por medio de diversas expresiones regulares. A diferencia del ejercicio realizado en clases, decidimos no utilizar la función *split* para separar en oraciones cada línea, porque separaba la información que debíamos encontrar.

Las búsqueda de cada dato está en una función que los imprime en “*salida.txt*” y los devuelve para que puedan ser guardados en una tabla en la base de datos en Postgres para tener un respaldo.

Se utilizó la librería *utf8* para que se puedan leer e interpretar acentos y eñes dentro del código. Así como también se utilizó el *encoding(UTF-8)* al momento de leer el dataset para que pueda ser leído sin cambiar los caracteres compuestos.

También se utilizó la librería *DBI* que es la encargada de realizar la conexión con *PostgreSQL* y escribir los datos en la tabla.



The screenshot shows a PostgreSQL data editor window titled "Edit Data - PostgreSQL 9.3 (localhost:5432) - postgres - estudiantes". The window displays a table with 25 rows and 12 columns. The columns are: id, nombre, edad, estatura, peso, correo, lugar_nacimiento, estado_civil, telefono, universidad, carrera, and carrera. The data is as follows:

	id	nombre	edad	estatura	peso	correo	lugar_nacimiento	estado_civil	telefono	universidad	carrera
	integer	character varying(100)	integer	character varying(100)	character varying(100)	character varying(100)	character varying(100)	character varying(100)	character varying(100)	character varying(100)	character varying(100)
1	1	Alejandro	28	1.88 metros	85 kilogramos	abenitez@es		casado	+593 986889		
2	2	Jorge	24								Ingeniería en Ciencias Computacionales especialización en Sistemas Multi
3	3	Erick	24	1.60 metros	55 kilogramos	santiago_6@			0991688888	ESPOL	Ingeniería en Sistemas Computacionales especialización Sistemas de Infor
4	4	Joe	23	1.69 metros		ure@espol.e			0989786378		Ingeniería Ciencias Computacionales
5	5	Victor	24				Guayaquil			ESPOL	Ingeniería en Ciencias Computacionales Orientación Tecnológico
6	6	Roberto	21	1.70 metros		tito_yc@hot			0989572020	ESPOL	
7	7	Ray	21	1.75 metros	70 kilogramos				0123456789	ESPOL	Ingeniería en Ciencias Computacionales
8	8	Sara Ramirez	23	1.59 metros	61 kilogramos	revrqq@hotm			0987902159	ESPOL	Ingeniería en Ciencias Computacionales Especialización en Sistemas Mult
9	9	Leonardo	19						0927491232	ESPOL	
10	10	Juan José	21			fasddr@espo					Ingeniería en Ciencias Computacionales
11	11	José Luis	22	1.63 metros	65 kilogramos	jlluis@hotm	Babahoyo		0987733337	ESPOL	Ingeniería en Ciencias Computacionales especialización Sistemas Multimed
12	12		28			fabo2185@no		soltero		ESPOL	Ingeniería en computación en la carrera de Sistemas de Multimedia
13	13	Dimitri	23			diaz_123@e	Guayaquil		2242266	ESPOL	Ingeniería En Ciencias Computacionales Orientación Sistemas Multimedia
14	14	Cristian Pi	20			orpiaco@esp					
15	15	Luis Eduard		1.69 metros	63 kilogramos	luis@hotmail			0999999999	ESPOL	
16	16	Janna Barre	21	1.60 metros	55 kilogramos					ESPOL	Ingeniería en Ciencias Computacionales especialización Sistemas de Infor
17	17	Ángel Guale	20			xngelgual@				ESPOL	Ingeniería en Ciencias Computacionales especialización Sistemas Tecnológ
18	18	Javier	23	1.70 metros		javierron12			''	ESPOL	Ingeniería en Ciencias Computacionales
19	19	Cindy	21							ESPOL	Ingeniería en Computación especialización Sistemas Multimedia
20	20		23	1.59 metros	54 kilogramos	jacocam@hotmail	Loja		0988888888	ESPOL	Ingeniería en Ciencias Computacionales
21	21	Diego	23			lp_0790@hotmail			0997031917	ESPOL	Ingeniería en Ciencias Computacionales
22	22		24	1.63 metros		abodef@gmail	Guayaquil		0998765432	ESPOL	Ingeniería en Ciencias Computacionales
23	23	Julio Realp	20			julio_realp			0934564354		Ingeniería en Ciencias Computacionales de la FIEC
24	24	Eddie Hurta	22	1.61 metros		jashdka@esp			0987777777	ESPOL	Ingeniería en Computación
25	25	Marco Sanch	23	1.67 metros	73 kilogramos	msachez@esp	Guayaquil		0984348877	ESPOL	Ingeniería en Ciencias Computacionales Orientación Sistemas de Informaci

25 rows.

Información parseada a una tabla en PostgreSQL

```
C:\Users\juse\Documents\GitHub\ProyectoPerl2014\salida.txt - Sublime Text 2 (UNREGISTERED)
File Edit Selection Find View Goto Tools Project Preferences Help

salida.txt
1 Nombre: Alejandro Benitez
2 Edad: 28
3 E-mail: abenitez@espol.edu.ec
4 Estatura: 1.88 metros
5 Peso: 85 kilogramos
6 Telefono: +593 98688922
7 Estado Civil: casado
8
9
10 Nombre: Jorge
11 Edad: 24
12 Carrera: Ingeniería en Ciencias Computacionales especialización en Sistemas Multimedia
13
14
15 Nombre: Erick
16 Edad: 24
17 Carrera: Ingeniería en Sistemas Computacionales especialización Sistemas de Información
18 Universidad: ESPOL
19 E-mail: santiago_g@mail.com
20 Estatura: 1.68 metros
21 Peso: 55 kilogramos
22 Telefono: 0991688888
23
24
25 Nombre: Joe
26 Edad: 23
27 Carrera: Ingeniería Ciencias Computacionales
28 E-mail: ure@espol.edu.ec
29 Estatura: 1.69 metros
30 Telefono: 0989786378
31
32
33 Nombre: Victor
34 Edad: 24
35 Carrera: Ingeniería en Ciencias Computacionales Orientación Tecnológico
36 Universidad: ESPOL
37 Ciudad de nacimiento: Guayaquil
38
39
40 Nombre: Roberto
41 Edad: 21
42 Universidad: ESPOL
43 E-mail: tito_yc@hotmail.com
44 Estatura: 1.78 metros
45 Telefono: 0989572828
46
47
48 Nombre: Ray
49 Edad: 21
50
Line 8, Column 1 Tab Size: 4 Plain Text
```

Salida en archivo txt

Bibliografía:

- http://es.wikibooks.org/wiki/Programaci%C3%B3n_en_P Perl/Introducci%C3%B3n
- <http://www.techrepublic.com/article/retrieve-data-from-a-postgresql-database-using-the-perl-dbi/>
- <http://www.postgresql.org/docs/8.2/static/sql-insert.html>
- http://perlenespanol.com/tutoriales/funciones/crea_tus_propias_funciones.html
- http://perlenespanol.com/tutoriales/funciones/funcion_split_y_join.html