

Proyecto de curso Valor 40 % Paradigmas 2017 Universidad Nacional

Eddy Miguel Ramírez Jiménez

1. Introducción

El mundo de los lenguajes de programación ha estado en constante evolución, cada uno con sus particularidades, si son compilados, interpretados o ambos (en una máquina virtual). Además se cuenta con muchas características que tienden a facilitar o entorpecer la creación de diferentes tareas.

Precisamente esto ha provocado que ningún lenguaje haya caído totalmente en desuso y a pesar de eso, se siguen creando lenguajes nuevos cada día, que sean adaptables a la solución de problemas.

Este proyecto tiene que ser desarrollado en varios módulos buscando que el estudiante se familiarice con los diferentes paradigmas y las características que tienen los lenguajes Erlang, Prolog, Java y eventualmente C.

2. Objetivos

1. Hacer análisis sobre una red social como Twitter
2. Hacer análisis estadístico sobre los enunciados de ciertas personas en la red social
3. Creación de estructuras de datos apropiadas para tales fines
4. Generar un informe dinámico con interfaz gráfica de los resultados.

3. Descripción del problema

Actualmente los tweets son mensajes de menos de 141 caracteres donde se pueden expresar cualquier cosa, incluso algunas incoherencias. Pero lo

interesante es que también se pueden hacer uso de los llamados *hashtag* o hacer mención de personas a través de su cuenta de Twitter (ejemplo: @realDonaldTrump) de modo que se generan diversos temas que se pueden indagar.

Se debe de hacer un programa que busque en los últimos n Tweets de una persona a quien llamaremos AMIGO buscando por:

1. Menciones de personas
2. Menciones de *hashtags*
3. URLs
4. Texto en general

Una vez que se tienen los últimos n tweets catalogados, se deben de hacer el mismo proceso con cada una de las personas mencionadas (simultáneamente) y se van a escanear sus n últimos tweets.

También se deben revisar los últimos n *hashtags* buscando el texto en general. A la vez que se va a revisar las páginas web ubicadas en los tweets y los *hashtags* presentes y se va a recopilar el texto de esas páginas.

Finalmente se va a buscar las k palabras más utilizadas por AMIGO en esos n tweets donde se va a hacer un control cruzado con las palabras más utilizadas por las personas a las que ha mencionado y con las páginas que han sido citadas o a través de *hashtags* (evitando artículos y preposiciones).

Para ello se van a generar una serie de diagramas de Venn donde se verían las palabras en común, tanto por AMIGO, sus contactos y los *hashtags*, donde la intensidad (o frecuencia) de esas palabras serán representadas por su tonalidad, cuánto más rojo es más frecuente, luego si es menos mencionada, sería representada en azul.

4. Descripción de los módulos

El programa anteriormente descrito debe de realizarse en varios módulos. El primero debe de ser el encargado de navegar en los tweets y debe de trabajar por hilos para poder hacer la navegación distribuida entre páginas web, *hashtags* y contactos mencionados, mediante la generación de procesos para hacer el escaneo simultáneamente.

Luego debe de pasar por un programa que revise mediante un trie los textos y genere un documento en un formato cualquiera para ser entregado al último módulo. Este archivo debe tener el conteo de las frecuencias de las k palabras más mencionadas en cada grupo.

Finalmente el último módulo debe de recibir ese archivo y generar un digrama de Venn que involucre las palabras más usadas de AMIGO (en cuadro verde), de sus contactos (en cuadro amarillo), de las páginas web suyas y de sus amigos (en cuadro blanco) y las de los hashtags y sus links (en cuadro negro). Respetando la intersección de los conjuntos y la frecuencia de las palabras.

El primer módulo debe estar hecho en Erlang, el segundo en Prolog y el tercero en Java.

Finalmente un programa en C recibirá el nick del AMIGO, el número de tweets que debe buscar (n) y el número de palabras más frecuentes que debe de mostrar (k). Este programa en C debe de ejecutar los demás módulos de forma transparente al usuario.

Cada grupo de trabajo debe de solicitar al profesor 4 personas que serán el AMIGO.

5. Evaluación

El proyecto debe de ser realizado en a lo sumo grupos de dos personas.

Cada uno de los módulos tiene un valor de 13 % de la nota final del curso y deben funcionar correctamente en Linux.

Los aspectos que serán valorados serán funcionalidad, respeto del paradigma y de las normas del lenguaje.

Es imperativo que todo programa cuenta con documentación interna que indique lo que hace.

El microprograma de C tiene un valor de 1 % de la nota final. La fecha de entrega es el lunes de semana 16.