

1. Lector básico de archivos de texto: revisar el código facilitado para la lectura de un archivo de texto: "processHTML.c".

La clase se encarga de abrir el archivo indicado (en caso de indicar uno), y tomar su contenido línea por línea. Esta línea es enviada al método processLine, el cual trabaja con expresiones regulares (para poder expresar patrones que coincidan con secuencias de caracteres). Estas expresiones se remueven del texto, quedando solo las palabras del mismo. Al mismo tiempo, mediante strtok se "divide" ese texto dependiendo del divisor especificado, logrando tomar solo palabra por palabra. Cada una de ellas se imprime en consola. El método culmina su tarea cuando el archivo finaliza.

2. Defina su estructura de datos para acumular la cantidad de etiquetas HTML que encuentra.

La estructura utilizada es un mapa. La llave de cada elemento es la etiqueta respectiva (un string), y el valor, corresponde al contador (un int). Así, cada vez que se compara las palabras del archivo (con cada etiqueta HTML posible) si alguna de estas coincide, se aumenta su propio contador. Además, la estructura es segura para la implementación de hilos, ya que los elementos en un mapa son estables, no se mueven o invalidan a menos que el elemento se borre del mapa.

3. Construya una clase "parserHTML" en C++ que reciba una línea y sea capaz de devolver, por medio de un iterador, todas las etiquetas HTML que existen en esa línea y la cantidad de veces que aparece.

Para la inicialización del mapa, se crea un método "inicializar", el cual agrega todas las etiquetas HTML posibles, declarandolas en 0. Estos serán sus contadores. Dentro del método processLine, facilitado por el profesor, se llama al método "comparar", cuando el token ya está preparado. Dentro de "comparar" se hace la comparación del token (palabra) ingresado con cada una de las llaves del mapa. Si alguna coincide, se aumenta el contador de la misma. Al finalizar el recorrido, el nuevo mapa actualizado con sus contadores se retorna. Al igual que sucede al finalizar el método de lectura del archivo. Para culminar, se llama al método "imprimir" el cual recorre el mapa de etiquetas, si alguna presenta un contador mayor que cero, significa que su llave (la etiqueta) apareció al menos una vez, así que se imprime en consola la etiqueta con su contador.

Esta clase se encuentra en la carpeta "Parser" de Semana7. Para la compilación del programa utilizar el makefile. Para su ejecución, hacerlo con el archivo HTML provisto en la carpeta: ejemplo.html.

```
jennifer@jennifer-VirtualBox: ~/Documentos/C++/Parser
Archivo Editar Ver Buscar Terminal Ayuda
jennifer@jennifer-VirtualBox:~/Documentos/C++/Parser$ make
g++ -g parserHTML.o -o parser
jennifer@jennifer-VirtualBox:~/Documentos/C++/Parser$ ./parser ejemplo.html
Lectura del archivo completada.
Etiqueta: !DOCTYPE. Cantidad de veces encontrada: 1.
Etiqueta: /body. Cantidad de veces encontrada: 1.
Etiqueta: /head. Cantidad de veces encontrada: 1.
Etiqueta: /html. Cantidad de veces encontrada: 2.
Etiqueta: /title. Cantidad de veces encontrada: 1.
Etiqueta: body. Cantidad de veces encontrada: 1.
Etiqueta: head. Cantidad de veces encontrada: 1.
Etiqueta: html. Cantidad de veces encontrada: 2.
Etiqueta: title. Cantidad de veces encontrada: 1.
jennifer@jennifer-VirtualBox:~/Documentos/C++/Parser$
```

(a) Ejecución del programa con ejemplo.html.