

# Trabajo práctico integrador

## Procesador de subtítulos

Seminario de lenguajes opción C

2018

El trabajo de este año consiste en el desarrollo en C de una aplicación que manipule archivos de subtítulos en formato SubRip.

El formato SubRip es uno de los formatos de subtítulos para vídeos más difundidos y compatibles. Normalmente los archivos SubRip tienen extensión .srt.

### 1. El formato

Un archivo SubRip está compuesto de una secuencia de subtítulos, cada subtítulo se almacena con el formato:

Figura 1: Formato SubRip

```
índice
tiempo_inicio --> tiempo_fin
Texto del subtítulo (una o más líneas)
línea_en_blanco
```

El índice es un número entero positivo representando el orden de los subtítulos, así el primer subtítulo siempre tendrá índice 1 y el tiempo se representa con formato HH:MM:SS,mmm donde:

1. HH son las horas relativas al comienzo del vídeo.
2. MM los minutos.
3. SS los segundos.
4. mmm las milésimas de segundo.

Figura 2: Ejemplo de subtítulo

1
00:00:20,000 --> 00:00:24,400
Altocumulus clouds occur between six thousand
2
00:00:24,600 --> 00:00:27,800
and twenty thousand feet above ground level.

## 2. Modalidad de la entrega

El trabajo se debe subir por <http://catedras.info.unlp.edu.ar> antes de las 23:55 del día de entrega establecido. Se puede realizar en grupos de 2 personas o individualmente.

Todo lo entregado se someterá a un testeo de similitudes usando el similarity tester de Dick Grune<sup>1</sup> seguido de una inspección visual de los casos sospechosos para detectar posibles copias, cualquier copia detectada es motivo suficiente para que los involucrados desapruében la cursada.

La reentrega de la etapa solamente está disponible para quienes entreguen en la primera fecha de entrega y asistan al coloquio correspondiente (desaprobando por errores o porque el código está incompleto y asistiendo al coloquio).

El calendario de entregas para lo que resta de la cursada es el siguiente:

Fecha	Instancia de evaluación
Semana del aa/bb	Entrega + coloquios.
Semana del cc/dd	Recuperatorio de la etapa 2 y coloquio de quienes reentregaron.

### 2.1. Entrega

Implementación de la funcionalidad completa de la aplicación como se detalla más abajo.

La entrega se considerará aprobada si se cumplen las siguientes condiciones:

- Se realiza la entrega en tiempo y forma.
- El programa funciona como es especificado en este documento sin fallar con ningún subtítulo de prueba ni combinación de parámetros (excepto que venga más de un

<sup>1</sup>[http://dickgrune.com/Programs/similarity\\_tester/](http://dickgrune.com/Programs/similarity_tester/)

*-f* o *-o*).

- El programa permite cargar archivos con líneas arbitrariamente largas sin fallar.
- El alumno asiste al coloquio y demuestra conocimiento sobre el código desarrollado.
- Se libera toda la memoria alguna vez alocada.
- Se lee correctamente desde archivos.
- Se hace un manejo correcto de arreglos y punteros.
- El programa está correctamente modularizado y usando correctamente los `.h`.
- El programa es legible por cualquier desarrollador y compila sin warnings.

### 2.1.1. Funcionalidad

El programa deberá permitir manipular archivos SubRip. Para ello aceptará los siguientes parámetros:

- *-f filein*  $\leftarrow$  *file* es el archivo a procesar.
- *-o fileout*  $\leftarrow$  *fileout* es el archivo donde se guardarán los cambios.
- *-b index*  $\leftarrow$  borra el subtítulo número *index*.
- *-i startmilis endmillis texto*  $\leftarrow$  inserta el subtítulo *texto* en *startmilis* con duración hasta *endmillis*, corrige todos los índices de los subtítulos siguientes.
- *-v*  $\leftarrow$  valida el subtítulo buscando los siguientes errores:
  - Los números de índice dentro del archivo no son números consecutivos ordenados.
  - Los números de índice no arrancan en 1.
  - Los subtítulos están solapados (el tiempo de fin del subtítulo *i* es mayor o igual al tiempo de inicio del subtítulo *i* + 1).
  - Dura menos de 1 segundo.
  - Dura más de 7 segundos.
  - Tiene líneas demasiado largas (más de 36 caracteres).
  - Tiene demasiados caracteres por segundo (más de 25).
  - Tiene más de 2 líneas.
  - Hay menos de 75 milisegundos entre el final del subtítulo *i* y el principio del *i* + 1.

Si la validación encuentra algún error programa deberá imprimir el número de índice del subtítulo y una descripción del error, y deberá seguir validando el archivo en busca de más errores. Luego deberá seguir procesando el resto de los parámetros por más que alguna validación falle.

Modo de procesar los parámetros:

- El parámetro *-f* es obligatorio y no se puede repetir.
- El parámetro *-o* es obligatorio cuando hay opciones que modifican al archivo (todas excepto *-v*) y no se puede repetir.
- El resto de los parámetros son opcionales, pueden repetirse y las acciones que representan deben ejecutarse en orden.
- Si un parámetro se repite, la acción indicada por el mismo deberá repetirse.
- Los parámetros se reciben en cualquier orden (es conveniente hacer 2 pasadas, en la primera se procesan *-f* y *-o*, y en la segunda se procesa el resto de los parámetros).
- Si se recibe más de un *-f* o *-o* el programa deberá mostrar un mensaje de ayuda con los parámetros válidos y terminar con código de salida 1.
- Si no se recibe un parámetro obligatorio deberá mostrar la ayuda anterior y terminar con código de salida 2.
- Si se recibe un parámetro inválido (por ejemplo: *-z* o *-cM*) deberá mostrar la ayuda anterior y terminar con código de salida 3.

Si no se puede abrir el archivo de entrada o el de salida, el programa deberá mostrar un error indicando el nombre del archivo que no pudo abrir y terminar con código de salida 4.

Por ejemplo la siguiente invocación al programa verifica los errores, luego agrega una entrada a las 7000 milisegundos (que se mostrará hasta las 8200 milisegundos) con el texto *.^ la grande le puse Cuca*, verifica el resultado, borra el subtítulo 3 y verifica el resultado:

---

```
./tpintegrador -v -f "los_simpsons.srt" -o "salida.srt" -i 7000 8200 "A la grande le puse Cuca" -v -b 3 -v
```

---

Subtítulos para probar: <http://www.subdivx.com/X2X-subtitulos.html>

**Nota 1:** Se puede usar cualquier funcionalidad incluida en GNU libc.

**Nota 2:** Oficialmente la película más larga de la historia “La cura para el insomnio”<sup>2</sup> dura 5.220 minutos, unos 313.200.000 milisegundos, tener esto en cuenta a la hora de almacenar los tiempos y duraciones.

---

<sup>2</sup>[http://es.wikipedia.org/wiki/The\\_Cure\\_for\\_Insomnia](http://es.wikipedia.org/wiki/The_Cure_for_Insomnia)