PRÁCTICAS DE LA ASIGNATURA
ANÁLISIS Y DISEÑO DE ALGORITMOS
GRADO EN INGENIERÍA INFORMÁTICA (UVa)

CURSO 2017/2018

ENUNCIADO PARA LA PRÁCTICA 2

Introducción

Las secuencias biológicas se pueden representar como secuencias de símbolos. Por ejemplo, las secuencias de ADN se pueden representar como secuencias de cuatro letras, *A*, *C*, *G* y *T*, que representan las cuatro bases que componen las cadenas de ADN (*Adenina*, *Citosina*, *Guanina* y *Timina*). Cuando se encuentra una nueva secuencia, es de interés descubrir a qué secuencias ya conocidas se parece más. Esto ayuda, entre otros, a descubrir la relación entre los genes responsables de enfermedades de tipo cancerígeno y aquellos que han tenido una evolución normal.

Lo que se busca en estos casos es la **subsecuencia** (o subsecuencias) de mayor longitud que aparece en todas las secuencias que se comparan. Es importante tener en cuenta que los símbolos o caracteres de las subsecuencias no tienen por qué ser contiguos. Por ejemplo, *ACE* es una subsecuencia de *ABCDE*.

Ejemplo:

Dadas las dos secuencias *S1* y *S2*, una subsecuencia de mayor longitud que aparece en ambas es *GCCAG* (Atención: no es necesariamente la única).

S1: *GCCCTAGCG*

S2: *GCGCAATG*

Objetivo

El objetivo de la práctica es obtener un algoritmo para obtener la subsecuencia de mayor longitud que aparece en dos secuencias de entrada. El algoritmo debe enmarcarse en alguna de las técnicas de diseño de algoritmos vistas en el tema 2 (algoritmos voraces, divide y vencerás, programación dinámica, ...). El algoritmo tomará como entrada dos secuencias, expresadas como dos cadenas de caracteres. Se considera válido cualquier carácter del abecedario castellano. Su salida será el conjunto de subsecuencias de mayor longitud que aparecen en ambas. Como mínimo debe resolverse el problema para encontrar una de estas subcadenas.

El programa que invocará el algoritmo tomará como entrada un fichero de texto que contendrá una pareja de secuencias, que son las que se comparan. Este programa generará como salida otro fichero de texto que contendrá el conjunto de subsecuencias encontradas.

Desarrollo de la práctica

Formato de los ficheros

El fichero de entrada tendrá nombre *entrada.txt*. Cada secuencia estará en una línea. Por ejemplo, el siguiente sería un fichero de entrada al programa para el ejemplo de este enunciado:

entrada.txt

```
GCCCTAGCG
GCGCAATG
```

El fichero de salida tendrá el mismo formato (una subsecuencia en cada línea) y tomará nombre *salida_p2_login1_login2_login3.txt*, donde *login1, login2* y *login3* son los identificadores de los miembros del grupo.

Opción 1 (obligatoria)

Se resolverá el problema usando dos técnicas de programación diferentes (se sugiere fuerza bruta y otra que elegirá el grupo). Como mínimo, ambas soluciones deben resolver el mismo problema: comprobar si la cadena Z es subsecuencia de X. Sin embargo, al menos una propuesta debe ser capaz de encontrar también una subsecuencia de la mayor longitud posible que aparezca en las dos secuencias de entrada, X y Z.

Un ejemplo de fichero de salida para la entrada del ejemplo anterior es el siguiente

p2_login1_login2_login3.txt

```
GCCAG
```

Opción 2 (opcional)

Se define igual que la opción 1, pero el resultado serán todas las subsecuencias de mayor longitud. Si se resuelve la opción 2, no es necesario resolver de modo independiente la fase 1. En el caso del ejemplo anterior obtendríamos un fichero similar al siguiente:

p2_login1_login2_login3.txt

```
GCCAG
GCGCG
GCCTG
```

Normas de entrega

La práctica se realizará en **grupos de tres personas**. Se debe entregar el código fuente de los programas que se han creado para la realización de la práctica (se recomienda el uso del lenguaje Java aunque alternativamente se permite realizar en cualquier lenguaje de programación). Dichos programas deben compilar y ejecutar perfectamente en las máquinas de los laboratorios.

Además se entregará un documento en **formato PDF** donde se indique lo siguiente:

- Nombre de ambos alumnos,
- Una breve descripción de cómo se ha realizado la práctica: tipo de algoritmo utilizado, algoritmo, razones para elegirlo, ...
- Instrucciones para la ejecución del programa,
- Comparación entre la solución de fuerza bruta y la solución elegida por el grupo. Debe indicar como mínimo: cuál de las dos soluciones es mejor y razones que apoyan esta afirmación.

Fecha de entrega

La fecha límite de entrega es **el 10 de noviembre a las 23:55**. Los ficheros de la práctica se empaquetarán en un fichero zip cuyo nombre será: p2_login1_login2_login3.zip, donde login1, login2 y login3 son los nombre de usuario en el laboratorio de los alumnos que forman el grupo. El fichero zip se entregará a través del Aula Virtual (http://aulas.inf.uva.es).

La defensa de la práctica se realizará en las semanas siguientes a la entrega, en el laboratorio. **La defensa de la práctica es obligatoria.**