

LA VERDAD ESTÁ AHÍ AFUERA

Los científicos terrestres están analizando el ADN de unos restos orgánicos hallados en un meteorito. Como resultado de estos análisis, han concluido que: 1) las cadenas de ADN contienen los cuatro compuestos habituales: A, C, G y T; 2) el compuesto A siempre aparece una única vez al principio de la cadena; 3) los compuestos aparecen siguiendo ciertas reglas que asocian compuestos a patrones de compuestos, de manera que cuando se encuentra uno de estos patrones en la cadena, se puede asegurar cuál es el siguiente compuesto que aparece en ella. Un ejemplo podría ser CGTG -> C, que significa que siempre que aparecen consecutivamente los compuestos CGTG en una cadena de ADN, puede afirmarse que tras ellos aparece el compuesto C.

Apartado 1 (3 puntos)

Una vez se han identificado las reglas mencionadas, se quiere tener un programa capaz de generar una cadena de ADN de una longitud determinada.

Objetivo

Dado un conjunto de reglas que definen la formación de las cadenas de ADN, y una longitud l , reconstruir a partir del conjunto de reglas una cadena de ADN de longitud l .

Entrada

La entrada del programa será un fichero de texto (ASCII) llamado ADN1.IN. Este fichero contendrá en su primera línea un número entre 0 y 9999 que será la longitud de la cadena que se quiere generar (sin contar la A inicial). El resto de este fichero contendrá una regla en cada línea, hasta un máximo de 100 reglas. De esta forma, en cada línea aparecerá el patrón seguido de un punto y seguido del compuesto que deberá aparecer siempre detrás de ese patrón en la secuencia. Los patrones considerados tendrán una longitud máxima de 20 caracteres.

El conjunto de reglas será mínimo, esto es, no se puede dar una situación en que a partir de una subcadena se puedan aplicar dos reglas simultáneamente. Además, este conjunto será completo, esto es, en cualquier situación siempre se encontrará una regla que se pueda aplicar.

Salida

Un fichero de texto (ASCII) con la secuencia calculada a partir de los datos de entrada. Este fichero no deberá tener ningún retorno de carro ni ningún otro carácter aparte de los caracteres 'A', 'C', 'G' y 'T' de la cadena. El nombre de este fichero será ADN1.OUT

Ejemplo: Dado el siguiente fichero de entrada:

ADN1.IN	ADN1.OUT
12	ACCGGCCCGCCCG
A.C	
AC.C	
ACC.G	
ACCG.G	
CCC.G	
CCCG.C	
GC.C	
GCC.C	
GG.C	

Que contiene una A seguida de 12 caracteres.

Explicación

La generación de la salida a partir de la entrada se muestra paso a paso a continuación:

A - se sabe que las cadenas empiezan por A

AC - tras aplicar la primera regla

ACC - tras aplicar la segunda regla

ACCG - tras aplicar la tercera regla

ACCGG - tras aplicar la cuarta regla

ACCGGC - tras aplicar la novena regla

ACCGGCC - tras aplicar la séptima regla
ACCGGCCC - tras aplicar la octava regla
ACCGGCCCCG - tras aplicar la quinta regla
ACCGGCCCCG - tras aplicar la sexta regla
ACCGGCCCCGCC - tras aplicar la séptima regla
ACCGGCCCCGCC - tras aplicar la octava regla
ACCGGCCCCGCCG - tras aplicar la quinta regla

Apartado 2 (7 puntos)

Un sabotaje ha destruido la base de datos de reglas de generación de cadenas de ADN. Afortunadamente, las muestras de ADN extraterrestre se conservan. Los científicos quieren ahora un programa que calcule las reglas a partir de las cadenas de ADN existentes.

Objetivo

Dada una cadena de compuestos 'A', 'C', 'G' y 'T', que cumple que la única aparición de 'A' se da justamente al inicio de la cadena, realizar un programa que calcule las reglas de generación tal y como se han descrito en el apartado anterior. Esto es, el programa deberá calcular las reglas que definen en cada momento el siguiente símbolo en la cadena. De esta forma, el programa de este apartado deberá hacer la operación inversa al programa del primer apartado.

Entrada

La entrada del programa consiste en una secuencia de caracteres 'A', 'C', 'G' y 'T'. Esta secuencia se almacena en un fichero de texto (ASCII) llamado ADN2.IN. No aparecerá ningún retorno de carro ni ningún otro carácter en el fichero. La secuencia comenzará por una A (la única que puede aparecer en la cadena) y tendrá por lo menos un carácter más. Considerar que, como máximo, habrá que mirar los 20 caracteres anteriores en la secuencia para determinar el siguiente símbolo.

Salida

La salida del programa se almacenará en un fichero de texto (ASCII) llamado ADN2.OUT. Este fichero contendrá una línea por cada regla encontrada en la cadena de símbolos. De esta forma, en cada línea aparecerá el patrón seguido de un punto y seguido del símbolo que aparece siempre detrás de ese patrón en la cadena. Las reglas deberán ordenarse por orden alfabético ascendente. El conjunto de reglas generado deberá ser mínimo, esto es, dada una subcadena cualquiera, sólo una regla podrá ser aplicada para generar el siguiente símbolo. Dicho de otra forma, ninguna regla deberá ser sufijo de otra. Por ejemplo, la regla 'CC.G' es sufijo de la regla 'GCC.C' porque la cadena 'CC' es sufijo de la cadena 'GCC'. Esto significa que si la secuencia es, por ejemplo, '...GCC' las dos reglas anteriores podrían aplicarse, lo cual es incorrecto.

Ejemplo:

ADN2.IN	ADN2.OUT
ACCGTCCGTC	A.C AC.C CC.G G.T T.C TC.C

Estas reglas indican que tras la subcadena 'A' aparece una 'C' en la secuencia. Tras la subcadena 'AC' aparece una 'C' en la secuencia y así sucesivamente. Obsérvese como, por ejemplo, no aparece como patrón en la respuesta la subcadena 'CCG' que determina el símbolo 'T'. Esto es así porque es suficiente con incluir la subcadena 'G' que en todas las apariciones en la cadena siempre está seguida por una 'T'.