

# Problema P05: memorias de traducción

DLSI

## 1. Descripción

Algunos sistemas de traducción automática almacenan segmentos de texto (es decir, palabras o secuencias de palabras) y su correspondiente traducción como “nice try”  $\rightarrow$  “buen intento”. La traducción se realiza entonces substituyendo segmentos del texto origen por las traducciones guardadas.

Para realizar la traducción de una frase  $w_1w_2\dots w_N$  queremos dividirla en segmentos cuya traducción esté en la memoria de forma que el número de segmentos sea lo más pequeño posible. Supondremos que todas las palabras están en el diccionario y que por tanto, cualquier palabra aislada es siempre un segmentos válido.

## 2. Instrucciones

Escribe una clase P05 con una función `int best(String[] data)` que calcule el número mínimo de segmentos en que se puede dividir una frase  $w_1\dots w_N$ . La entrada consiste en cadenas de ceros y unos en las que el valor  $j$  de `data[i]` es 1 si  $w_i\dots w_j$  es un segmento almacenado y cero en caso contrario. Los elementos de la diagonal son siempre 1 y aquellos con  $i > j$  son siempre 0. Por supuesto, la longitud de cualquier cadena `data[n]` coincidirá con la dimensión del vector `data`.

Por ejemplo, si `data` es

```
{"1100000",  
 "0101000",  
 "0010100",  
 "0001000",  
 "0000101",  
 "0000010",  
 "0000001"}  
}
```

la salida es 3.