Estructuras de Datos (Gestión), curso 2003/04

Segunda Práctica

Introducción

Los diseñadores de teléfonos móviles se enfrentan al problema de conseguir que un usuario pueda introducir texto en un determinado lenguaje (español, inglés, chino, etc.) utilizando un teclado muy limitado (de 8 a 12 botones), lo que hace imprescindible asignar varios caracteres a cada tecla y definir un mecanismo para poder indicar cual es el carácter que queremos escribir de entre los asignados a cada tecla.

Si el alfabeto de un lenguaje consta de *n* caracteres y hay *m* teclas disponibles, la mayoría de los teléfonos móviles utilizan la estrategia de dividir el alfabeto (en su orden natural) en *m* bloques contiguos y asignar, en orden, cada bloque a una tecla. Cuando una tecla se pulsa una sola vez se introduce el primer carácter del bloque asignado a la tecla, si se pulsa dos veces se introduce el segundo carácter del bloque, si se pulsa tres veces el tercero, etc.

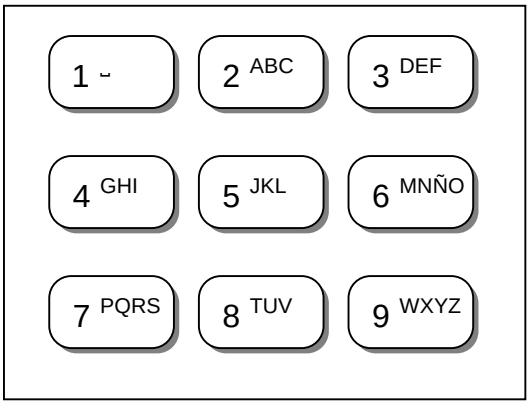

Fig. 1*: Teclado español*

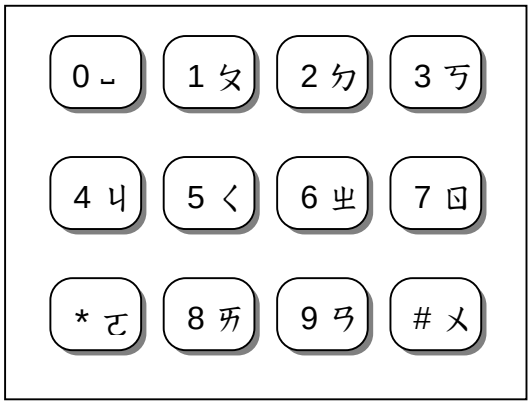

Fig. 2*: Teclado bopomofo*

En la Fig.1 se muestra la disposición típica de un teclado español, con 28 caracteres (el primero de ellos es el espacio en blanco, representado por el símbolo ␣) y 9 teclas disponibles: Para introducir el carácter P se debe pulsar una vez la tecla 7, para introducir el carácter S se debe pulsar cuatro veces la tecla 7.

En la Fig. 2 se muestra la disposición de un teclado para un lenguaje que utiliza el alfabeto bopomofo (se muestra en la Tabla 1), con 38 caracteres y 12 teclas disponibles, donde cada tecla sólo muestra el primer carácter del bloque (por ejemplo, para introducir el carácter ㄝ se debe pulsar tres veces la tecla * , que es la novena tecla disponible).

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
␣	ㄅ	ㄆ	ㄇ	ㄈ	ㄉ	ㄊ	ㄋ	ㄌ	ㄍ	ㄎ	ㄏ	ㄐ	ㄑ	ㄒ	ㄓ	ㄔ	ㄕ	ㄖ

20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
ㄗ	ㄘ	ㄙ	ㄚ	ㄛ	ㄜ	ㄝ	ㄞ	ㄟ	ㄠ	ㄡ	ㄢ	ㄣ	ㄤ	ㄥ	ㄦ	ㄧ	ㄨ	ㄩ

Tabla 1: *Alfabeto Bopomofo*

En todos los lenguajes existen algunos caracteres que aparecen con mayor frecuencia que otros (en español, por ejemplo, la letra *e* aparece con una frecuencia mucho mayor que la letra *x*). Si se analizan muchos textos escritos en un lenguaje determinado es posible obtener una tabla que indique la *frecuencia promedio* con la que aparece cada carácter en un texto cualquiera. La Tabla 2 muestra esas frecuencias (truncadas al tercer decimal) para el español.

1	2	3	4	5	6	7	8	9	10	11	12	13	14
_	A	B	C	D	E	F	G	H	I	J	K	L	M
0,16	0,098	0,01	0,043	0,048	0,108	0,006	0,01	0,003	0,065	0,003	0,000	0,05	0,02

15	16	17	18	19	20	21	22	23	24	25	26	27	28
N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
0,056	0,001	0,08	0,022	0,004	0,058	0,067	0,04	0,03	0,006	0,000	0,001	0,009	0,002

Tabla 2: *Probabilidad de aparición de cada carácter en un texto en español*

Conociendo éstas frecuencias es posible diseñar el teclado de un teléfono móvil para un determinado lenguaje de manera que se *minimicen* el número de pulsaciones necesarias para introducir texto. Si se multiplica la frecuencia de cada carácter por el número de pulsaciones necesarias para introducirlo y se suman todos esos productos se obtiene el promedio de pulsaciones para un texto cualquiera. El objetivo es elegir, de entre todas las formas posibles en que se puede dividir el alfabeto en bloques para asignar cada bloque a una tecla, la división que minimice ese promedio.

Requisitos de la Práctica

La práctica consistirá en crear un programa que reciba como entrada los valores de *n* (número de caracteres del alfabeto), *m* (número de teclas disponibles) y el vector f_1 ,..., f_n donde f_i indica la frecuencia promedio con que aparece el carácter *i*-ésimo del alfabeto en un texto (evidentemente, la suma de todas las frecuencias debe ser la unidad). El valor de *n* puede estar comprendido entre 1 y 300, y el de *m* entre 1 y 15.

El programa calculará la forma óptima de dividir el alfabeto en *m* bloques contiguos (cada bloque estará asociado a una tecla) de manera que se minimice el número promedio de pulsaciones necesarias para introducir texto en ese lenguaje, de acuerdo con las frecuencias de aparición de cada carácter indicadas por el vector ***f***.

El programa escribirá por pantalla esa solución (expresada mediante el vector descrito en los párrafos siguientes) y el valor promedio de pulsaciones que se obtiene con ella.

La forma de expresar la solución será utilizar un vector s_1 ,..., s_n con $s_i \in \{0,1\}$ y donde $s_i = 1$ indique que el carácter *i*-ésimo del alfabeto es el primero del bloque asignado a la tecla correspondiente y $s_i = 0$ indique lo contrario (el carácter *i*-ésimo no es el primero de su bloque). Evidentemente $s_1 = 1$ (el primer carácter del alfabeto debe ser el primero de su bloque) y el número de 1's que aparezcan en el vector ***s*** debe ser igual a *m*.

Por ejemplo, si la disposición del teclado de la Fig. 2 fuera óptima, el programa debería escribir por pantalla el vector ***s*** de la siguiente forma: (**Nota**: No es necesario que el programa muestre los índices del vector)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
1	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**1**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**1**	**0**

Calcular el número promedio de pulsaciones de una solución es muy sencillo: Si denominamos p_i al número de pulsaciones necesarias para introducir el carácter *i*-ésimo, entonces el promedio será la suma de los productos $p_i \times f_i$. Los valores p_i se calculan de la siguiente forma: Si $s_i = 1$ entonces $p_i = 1$ y si $s_i = 0$ entonces $p_i = p_{i-1} + 1$.

Por ejemplo, dada la disposición del teclado expresada en la Fig. 1 (que puede o no ser óptima) y la frecuencia de caracteres dada por la Tabla 2 se tendrían los siguientes valores para el vector ***s***, ***p*** y el promedio de pulsaciones:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
	␣	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
s	**1**	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**
p	**1**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**1**	**2**	**3**	**4**

promedio de pulsaciones = 2,135

Éste problema se puede resolver de muchas formas distintas, y se calificarán mejor aquellas soluciones que sean más eficientes. Se debe tener en cuenta que el valor de *m* está acotado, pero el valor de *n* puede ser bastante grande (existen lenguajes con un gran número de caracteres).

Documentación que se debe entregar

Se deberá entregar el código fuente del programa que resuelve la práctica (se recomienda el uso del lenguaje Eiffel, aunque alternativamente se permite realizarlo en Java o Delphi/Kylix) y un documento (impreso o en formato Word, Latex (DVI) o Acrobat PDF) donde se indique lo siguiente:

- Una breve descripción de la estrategia de diseño utilizada para resolver el problema planteado en la práctica.
- Una evaluación en notación asintótica de la eficiencia del programa (no es imprescindible que sea exacta: se pueden utilizar cotas inferiores y superiores).

La entrega de la práctica se realizará durante el examen teórico de la asignatura (mediante un disquete y opcionalmente documento impreso). La documentación de ésta práctica se puede adjuntar a la del resto de prácticas de la asignatura (no es necesario presentarla por separado).

Comprobación de los programas

En la página de la asignatura se publicarán casos de prueba para que podáis evaluar vuestras programas.