



CENTRO SUPERIOR DE INFORMÁTICA
Departamento de Estadística, I. O. y Computación
Teoría de Automatas y Lenguajes Formales

Listar todas las palabras de diferentes longitudes presentes en un fichero de texto

Realizar un programa en C que lea un fichero de texto y emita por pantalla una lista de las palabras que contiene ordenadas alfabéticamente e indicando el número de veces que han aparecido cada una de ellas. Para leer el texto se utilizarán las funciones de C para manipulación de ficheros de alto nivel (texto).

Las palabras podrán aparecer en el texto separadas entre sí por los caracteres espacio(' '), nueva línea ('\n'), tabulador ('\t'), salto de página ('\f'), retorno de carro ('\r'), coma (','), punto ('.'), punto y coma (;) o dos puntos (':'). Por tanto las palabras podrán contener cualquier carácter excepto alguno de los separadores anteriores. No se hará distinción entre letras mayúsculas y minúsculas.

Para permitir un acceso rápido a cada palabra, éstas deberán almacenarse en una estructura dinámica de tipo árbol binario ordenado, donde cada nodo almacenará la cadena correspondiente a la palabra y el número de veces que ha aparecido en el texto.

Para un fichero con el siguiente texto:

Si Q es nada, Q – nada es nada.

El árbol de búsqueda asociado sería el de la Figura 1.1.

Cada una de las líneas de la lista que deberá visualizarse por pantalla una vez que se haya leído todo el texto tendrá la siguiente estructura:

palabra : nº_veces_que_ha_aparecido

Así, para el ejemplo anterior el programa mostraría:

```
-      : 1
es     : 2
nada   : 3
si     : 1
q      : 2
```

El comportamiento del programa debería ser análogo al de los programas:

ftp://ftp.csi.ull.es/pub/asignas/AUTOMALF/p01_palabras/palabras

ftp://ftp.csi.ull.es/pub/asignas/AUTOMALF/p01_palabras/palabras.exe

versiones Unix y DOS respectivamente y accesibles en el servidor ftp del CSI.

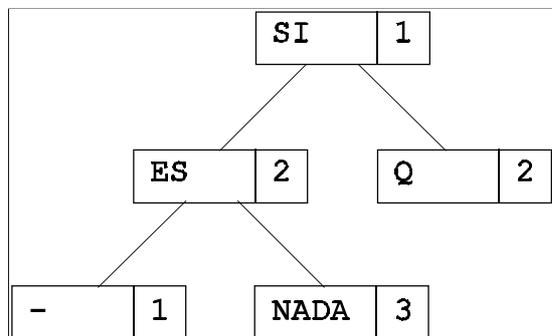


Figura 1.1 Arbol binario ordenado para el texto

Recomendamos consultar las referencias [Sed90], [Ten90] y [Wir80] para la realización de esta práctica.