

PRÁCTICA 4

Recuperación de la información



12 DE DICIEMBRE DEL 2018
VÍCTOR ANDRÉS VIZÁN

Contenido

1.	Introducción	2
2.	Desarrollo	3
3.	Pruebas.....	8
4.	Conclusiones y observaciones.....	10
5.	Bibliografía	11

1. Introducción

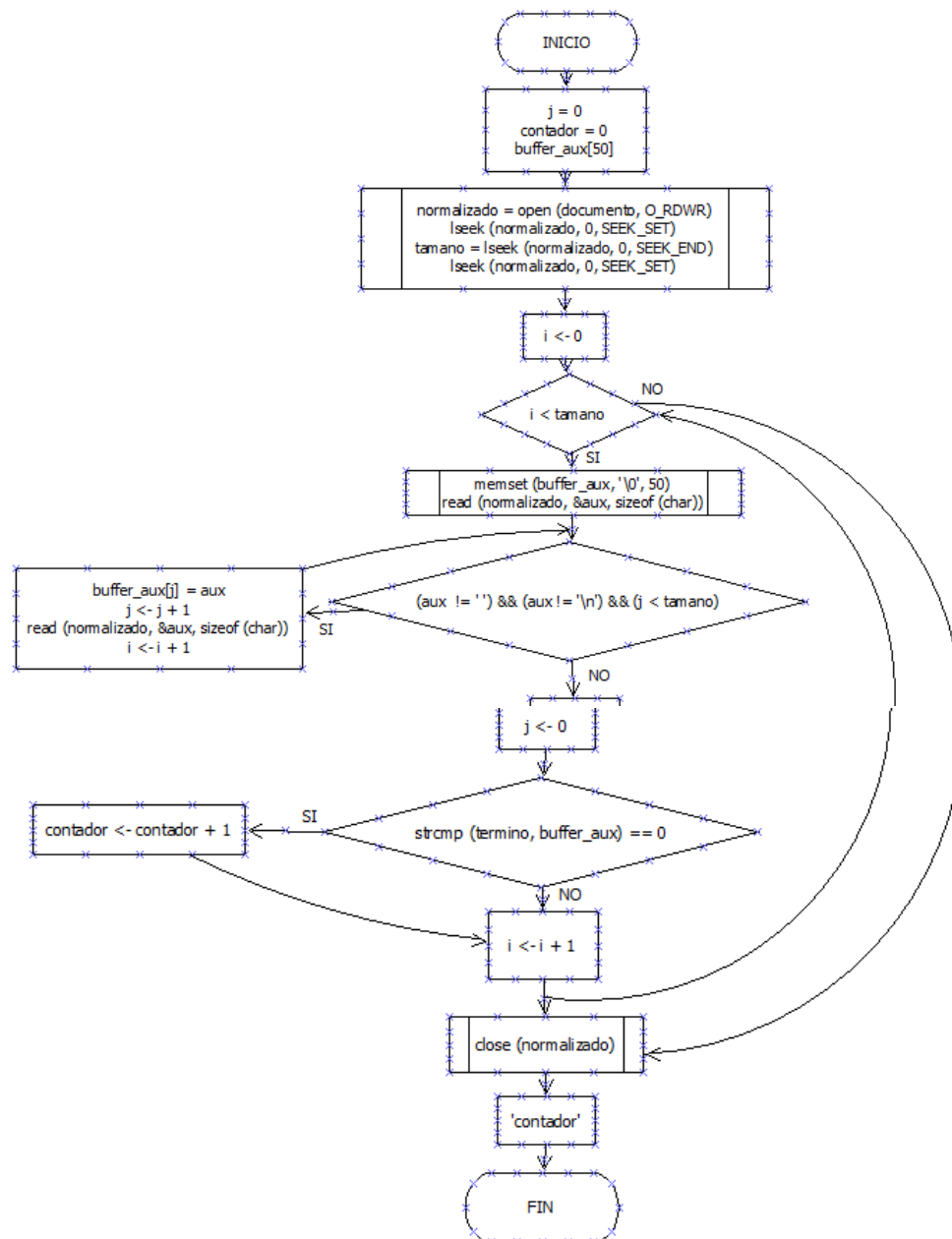
Esta cuarta práctica de la asignatura usa los códigos de la anterior práctica para hacer una consulta al programa mediante el método probabilístico.

El programa se estructura en diversos puntos:

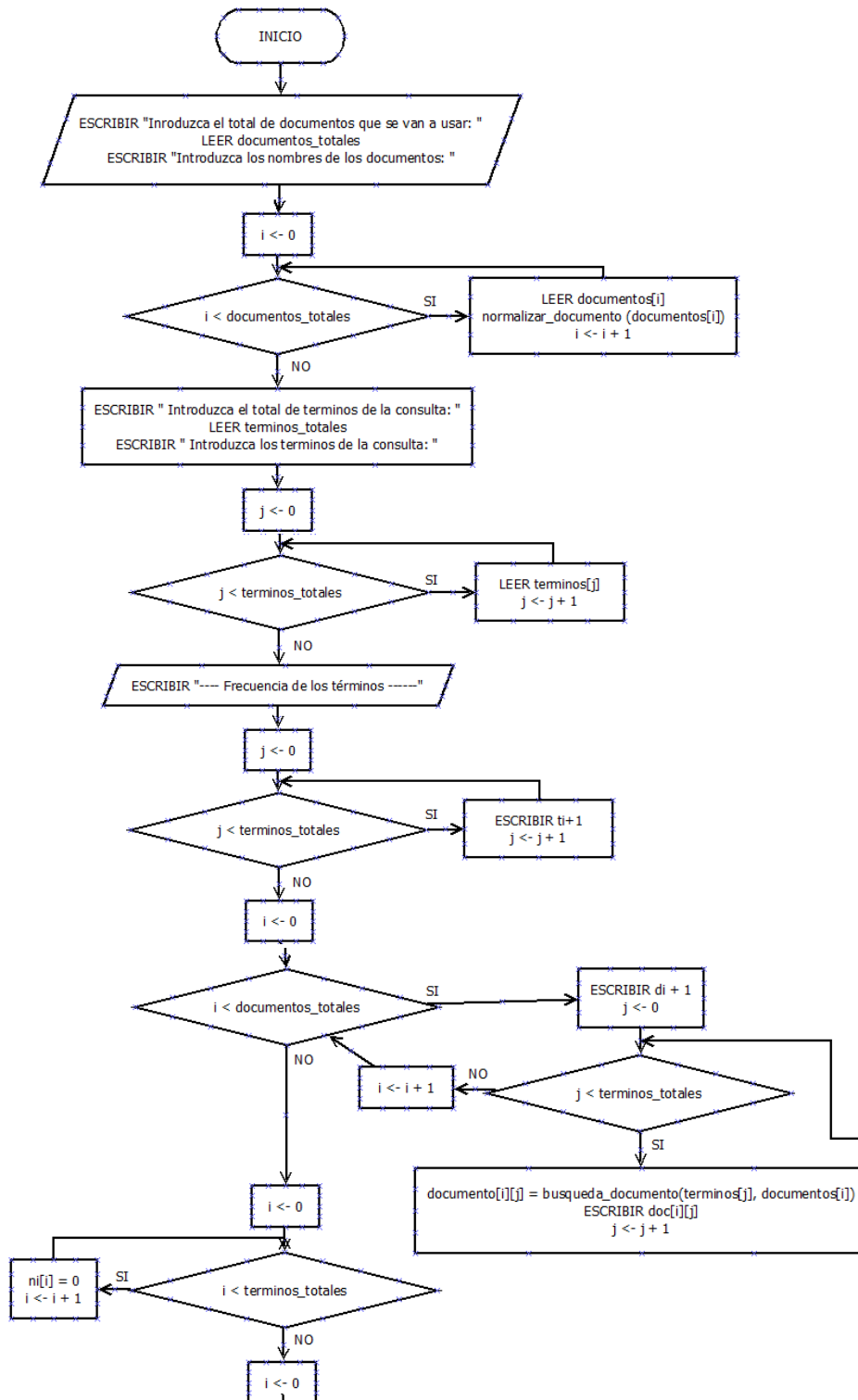
- Seguir un menú con diferentes opciones.
- La primera opción del menú te permite normalizar documento como en la práctica 1 y 2.
- La segunda opción te permite hacer una consulta mediante el método probabilístico introduciendo la consulta y el nombre de los documentos por teclado.
- La tercera opción te permite salir del programa.

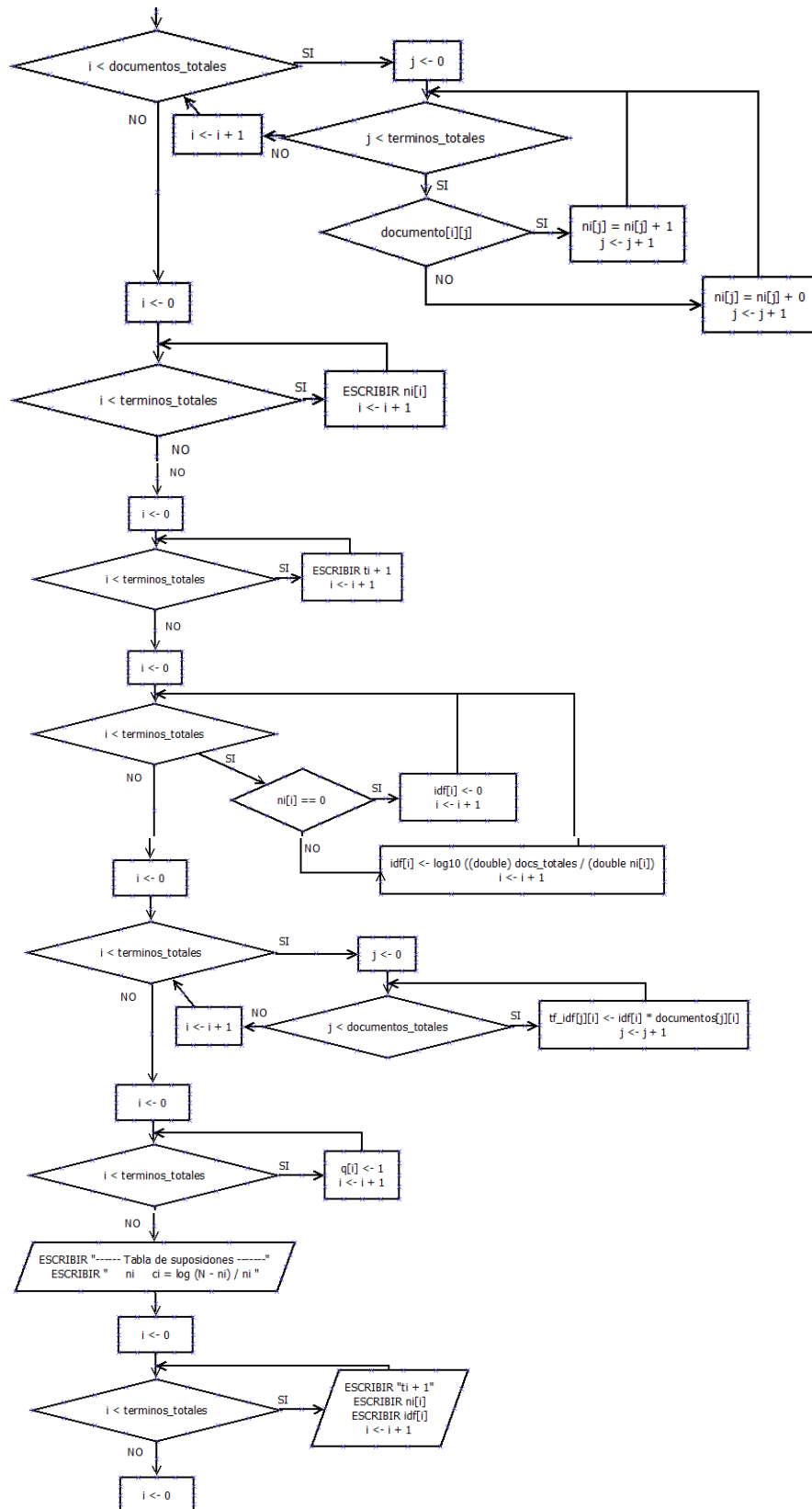
2. Desarrollo

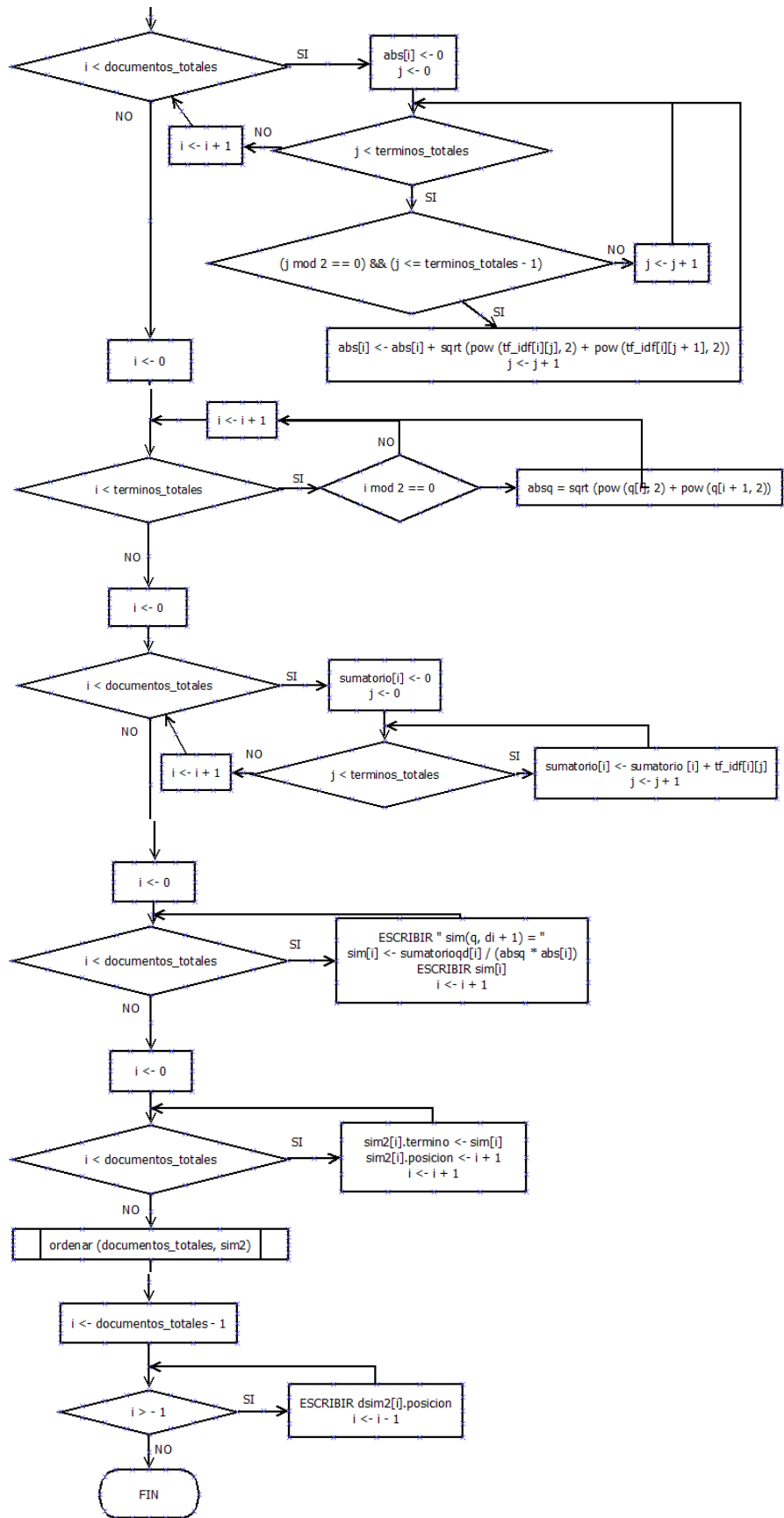
Función que busca un término introducido por parámetro en un documento (buscar_termino).



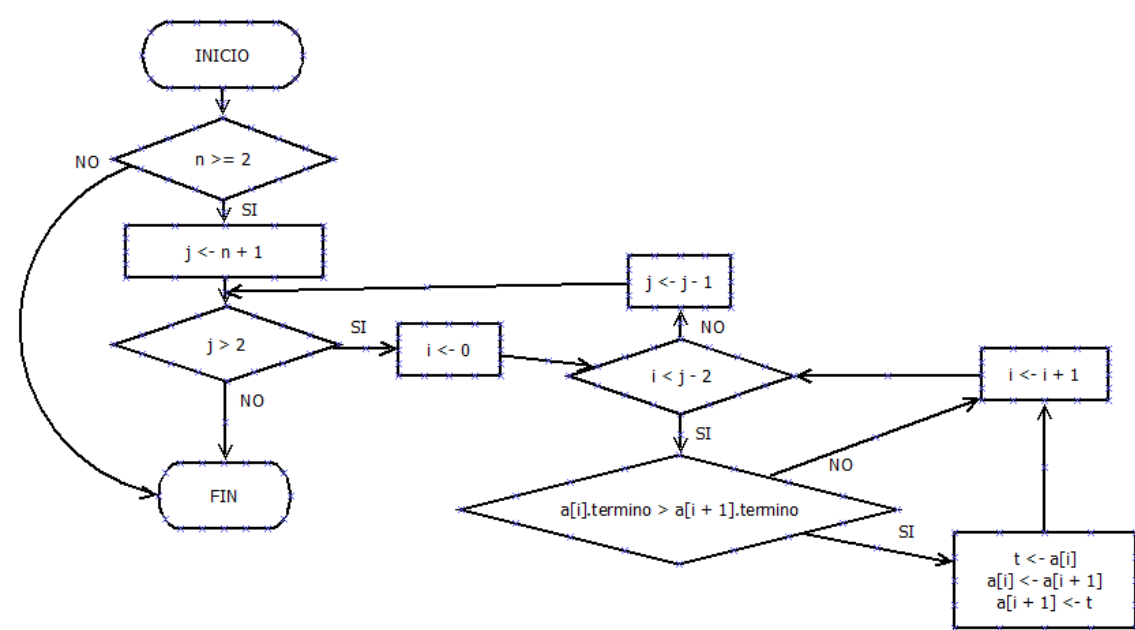
Módulo que hace una consulta.







Módulo que ordena un array de enteros.



3. Pruebas

Las siguientes capturas de pantalla muestran las diferentes opciones que el programa creado

En esta primera captura de pantalla se puede ver como al escoger la opción 1 del programa normaliza el documento introducido.

```
victor@Algoritmia:~/Escritorio$ gcc Practica4RI.c -lm
victor@Algoritmia:~/Escritorio$ ./a.out

Menu
1. Normalizar documento.
2. Hacer consulta.
3. Salir del programa.
Respuesta: 1

Indique el nombre del documento:
documento1.txt

Documento normalizado.
```

En estas tres capturas siguientes se muestra como introducir la información para que el programa haga la consulta. En la segunda imagen te muestra la información que el programa muestra. Y la tercera muestra el resultado de la consulta.

```
Menu
1. Normalizar documento.
2. Hacer consulta.
3. Salir del programa.
Respuesta: 2
Introduzca el total de documentos que se van a usar: 4
Introduzca los nombres de los documentos:
documento1.txt
documento2.txt
documento3.txt
documento4.txt
Introduzca el total de terminos de la consulta: 3
Introduzca los terminos de la consulta:
barco
casa
perro
```

```
----- Frecuencia de los terminos -----
      t1      t2      t3
d1      1      0      1
d2      0      1      1
d3      0      0      0
d4      1      1      1
ni      2      2      3

----- Tabla de suposiciones -----
      ni      ci = log (N - ni) / ni
t1      2      0.301030
t2      2      0.301030
t3      3      0.124939
```

```
sim(q,d1) = 0.000045
sim(q,d2) = 0.000045
sim(q,d3) = 0.000000
sim(q,d4) = 0.000060

Solucion: { d4 d2 d1 d3 }
```

Y en esta última captura de pantalla se muestra la tercera opción, que es la de salir del programa.

```
Menu
1. Normalizar documento.
2. Hacer consulta.
3. Salir del programa.
Respuesta: 3
victor@Algoritmia:~/Escritorio$
```

4. Conclusiones y observaciones

Casi todo el código es reutilizado de las prácticas anteriores. La diferencia está en el módulo consultar y en el de buscar_término. Esto hace la diferencia entre las prácticas, la segunda práctica el modelo booleano, la tercera el vectorial, y esta es el modelo probabilístico.

Se ha reutilizado casi todo el código de la primera, segunda y tercera práctica. Ha habido que cambiar un par de cosas al módulo normalizar_documento (), para introducir un fichero por parámetros.

Existe también en esta práctica un módulo para ordenar arrays de números.

Al final con estas prácticas además de seguir la asignatura, aprendes C y te pones a pensar en cómo hacer un programa de forma autodidacta.

5. Bibliografía

[1] "C library function tolower()"

https://www.tutorialspoint.com/c_standard_library/c_function_toupper.htm

[2] "Dia" <http://dia-installer.de/index.html.es>

[3] "C | Mi Word Press" <https://pasky.wordpress.com/category/programacion/c/>