

Memoria de la Práctica de Ficheros

Ficheros y Bases de Datos

ITIG

2008-2009

<i>Autores</i>		
<i>Apellidos</i>	<i>Nombre</i>	<i>NIA</i>

Índice

1. Introducción	2
2. Sistema con Organización Serial Consecutiva	3
2.1 Diseño Físico	3
2.1.1 Deficiencias del Diseño Físico del Fichero Original	3
2.2 Diseño Lógico	3
2.2.1 Deficiencias del Diseño Lógico del Fichero Original	3
2.3 Gestión del Espacio de Almacenamiento	4
2.4 Gestión del Tiempo de Acceso	4
3. Sistema con Organización Serial No Consecutiva	5
3.1 Diseño Optimizado del Registro	5
3.1.1 Diseño Lógico	5
3.1.2 Diseño Físico	5
3.2 Gestión del Espacio de Almacenamiento	6
3.3 Gestión del Tiempo de Acceso	6
3.4 Conclusiones de la Organización Serial No Consecutiva	6
4. Sistema con Organización Direccionada	7
4.1 Diseño del Cubo	7
4.2 Dimensionado del Fichero	7
4.3 Elección de la Función de Transformación	7
4.4 Gestión de Desbordamientos	7
4.5 Gestión del Espacio de Almacenamiento	8
4.6 Gestión del Tiempo de Acceso	8
4.7 Comparación con las Organizaciones Anteriores	8
5. Sistema con Organización Auxiliar: Índices	9
5.1 Gestión del Espacio de Almacenamiento	9
5.2 Gestión del Tiempo de Acceso	9
5.3 Comparación con las Organizaciones Anteriores	9
6. Conclusiones	10
7. Anexo 1: Ejemplo de Diseño Lógico de un Registro	11
8. Anexo 2: Ejemplo de Representación del Diseño Físico-Lógico de un Registro	12

Índice de Figuras

Figura 1. Diseño lógico del registro del sistema de ficheros	3
Figura 2. Diseño lógico del registro del sistema de ficheros	5

Índice de Tablas

Tabla 1. Diseño físico-lógico del registro del sistema de ficheros	3
Tabla 2. Espacio de almacenamiento con organización serial consecutiva	4
Tabla 3. Tiempos de acceso con organización serial consecutiva	4
Tabla 4. Diseño físico-lógico del registro del sistema de ficheros	5
Tabla 5. Tamaño del registro del sistema de ficheros	5
Tabla 6. Espacio de almacenamiento con organización serial no consecutiva	6
Tabla 7. Tiempos de acceso con organización serial no consecutiva	6
Tabla 8. Elementos de datos de control del cubo	7
Tabla 9. Estudio de la dispersión	7
Tabla 10. Espacio de almacenamiento con organización direccionada	8
Tabla 11. Tiempos de acceso con organización direccionada	8
Tabla 12. Índices implementados	9
Tabla 13. Espacio de almacenamiento con organización auxiliar indexada	9
Tabla 14. Tiempos de acceso con organización auxiliar indexada	9

1. Introducción

A lo largo de esta práctica se aplicarán las distintas organizaciones base estudiadas en clase para implementar un sistema de ficheros determinado. La realización de la práctica permitirá familiarizarse con ellas y comparar sus ventajas y desventajas.

La práctica consta de un conjunto de **apartados troncales que supondrán 7,5 puntos** de la nota de la práctica. Estos apartados deberán ser realizados de forma obligatoria por todos los alumnos y serán evaluados como sigue:

- 1) Sistema con Organización Serial Consecutiva (**hasta 1 punto**).
- 2) Sistema con Organización Serial No Consecutiva (**hasta 1,5 puntos**).
- 3) Sistema con Organización Direccionada (**hasta 2,5 puntos**).
- 4) Sistema con Organización Auxiliar: Índices (**hasta 1,5 puntos**).
- 5) Conclusiones (**hasta 1 punto**).

Para completar la nota de la práctica, se propone al alumno realizar mejoras sobre la versión básica de la práctica. Estas se evaluarán como sigue:

- 6) Optimización adicional de los registros. (**hasta 0,5 puntos**).
- 7) Implementación de la funcionalidad de inserción de nuevos registros en el fichero (**hasta 0,5 puntos**).
- 8) Navegación al registro anterior al mostrar los resultados de una consulta (**hasta 0,5 puntos**).
- 9) Estudio de otras funciones de dispersión (**hasta 1 punto**)
- 10) Otras estrategias para la gestión del desbordamiento (**hasta 1 punto**).
- 11) Implementación de más de dos índices, índices por esquemas de bits e índices arbóreos (**hasta 2,5 puntos**).

Comentarios adicionales:

1. Para la realización de la memoria de la práctica, deberá seguirse estrictamente esta plantilla, reemplazando las instrucciones que aquí aparecen por los resultados, explicaciones, argumentos y justificaciones requeridas en cada uno de los apartados.
2. Si fuese necesario entregar la memoria en formato electrónico, se mandará al correo del profesor de prácticas correspondiente en formato “pdf”.

2. Sistema con Organización Serial Consecutiva

En este apartado se describirá el diseño del fichero original ‘malos.txt’. Este diseño habrá sido obtenido por ingeniería inversa.

2.1 Diseño Físico

Se describirá el diseño físico de los registros del fichero original. Se detallará qué elementos de datos contiene un registro, cuál es su longitud y su codificación (o estructura cuando se trate de grupos repetitivos y vectores). Así mismo, se indicará la frecuencia de repetición de cada elemento de datos y su longitud (Ver Anexo 2: Ejemplo de Representación del Diseño Físico-Lógico de un Registro, más adelante).

Tabla 1. Diseño físico-lógico del registro del sistema de ficheros

Elemento de datos	Tipo	Codificación / Estructura	Frecuencia de repetición	Tamaño (bytes)

2.1.1 Deficiencias del Diseño Físico del Fichero Original

Se describirán las deficiencias encontradas a nivel físico en el fichero analizado, tales como codificaciones poco óptimas de determinados campos y otras causas de desperdicio del espacio, etc.

2.2 Diseño Lógico

Se detallará el diseño lógico del registro, mostrando la descripción y disposición de los elementos de datos y su tipo (grupos repetitivos, vectores, campos), así como su longitud máxima. También se reflejará si son o no repetitivos y obligatorios (Ver Anexo 1: Ejemplo de Diseño Lógico de un Registro, más adelante).

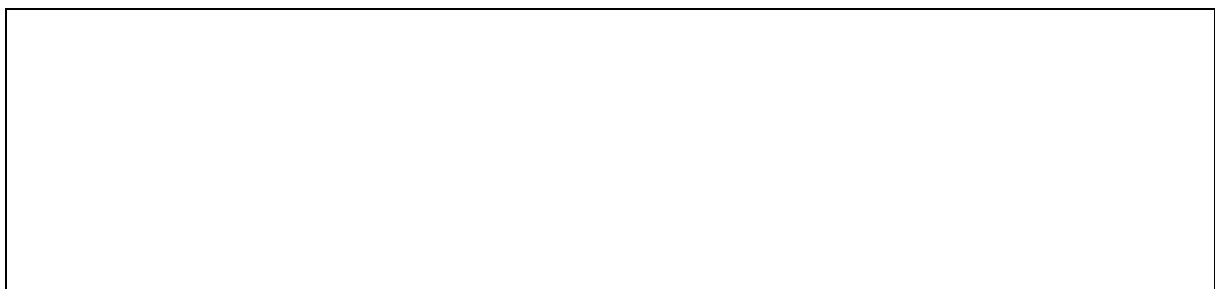


Figura 1. Diseño lógico del registro del sistema de ficheros

Se enumerarán los supuestos que se desprendan del análisis del fichero y que no aparezcan reflejados en el enunciado de la práctica (dominio de los campos enumerados, rangos de fechas que se considerarán, etc.).

2.2.1 Deficiencias del Diseño Lógico del Fichero Original

Se describirán las deficiencias encontradas a nivel lógico en el fichero analizado, tales como campos redundantes, campos a ser fusionados, campos a ser divididos, etc.

2.3 Gestión del Espacio de Almacenamiento

Rellenar la Tabla 2 con los valores del factor de bloqueo, densidad ideal y densidad real correspondientes al sistema de ficheros de organización base serial consecutiva. Adjuntar el desarrollo de los cálculos y explicar de forma razonada los resultados.

Tabla 2. Espacio de almacenamiento con organización serial consecutiva

Factor de bloqueo	
Densidad ideal	
Densidad real	

2.4 Gestión del Tiempo de Acceso

En este apartado se detallará, de forma teórica, el número mínimo, medio y máximo de accesos a memoria necesarios para los dos procesos de consulta descritos según el sistema de ficheros con organización serial consecutiva. Se justificarán los valores reflejados en la tabla para cada uno de los casos.

Tabla 3. Tiempos de acceso con organización serial consecutiva

Clave de consulta	Caso		
	Mejor	Medio	Peor
Seud			
Ape1			
Titu			

3. Sistema con Organización Serial No Consecutiva

Esta versión se centra en la optimización del diseño de los registros. Se describirá el nuevo registro del sistema de ficheros y se reflexionará sobre como repercuten estas mejoras en la gestión del espacio de almacenamiento y/o en el tiempo de acceso.

3.1 Diseño Optimizado del Registro

Se diseñará un nuevo registro para subsanar algunas de las deficiencias detectadas en los registros del fichero original. Se deberán eliminar, al menos, los espacios vacíos de los campos de longitud variable. El resto de optimizaciones serán evaluadas como mejoras sobre la versión básica.

3.1.1 Diseño Lógico

Se detallará el diseño lógico del registro, mostrando la descripción y disposición de los elementos de datos y su tipo (grupos repetitivos, vectores, campos), así como su longitud máxima. También se reflejará si son o no repetitivos y obligatorios (Ver Anexo 1: Ejemplo de Diseño Lógico de un Registro, más adelante).



Figura 2. Diseño lógico del registro del sistema de ficheros

Se enumerarán los supuestos tenidos en cuenta para el diseño, especialmente los implícitos, no reflejados en el enunciado de la memoria o considerados por el alumno.

3.1.2 Diseño Físico

Se describirá el diseño físico-lógico de los registros del fichero. Se detallará qué elementos de datos contiene, cuál es su longitud y su codificación (o estructura cuando se trate de grupos repetitivos y vectores). Así mismo, se indicará la frecuencia de repetición de cada elemento de datos y su longitud (Ver Anexo 2: Ejemplo de Representación del Diseño Físico-Lógico de un Registro, más adelante).

Tabla 4. Diseño físico-lógico del registro del sistema de ficheros

Elemento de datos	Tipo	Codificación / Estructura	Frecuencia de repetición	Tamaño (bytes)

Así mismo, se detallará el tamaño máximo teórico del registro y su tamaño medio obtenido experimentalmente:

Tabla 5. Tamaño del registro del sistema de ficheros

Tamaño del	Máximo	
------------	--------	--

registro	Medio	
----------	-------	--

3.2 Gestión del Espacio de Almacenamiento

Rellenar la Tabla 6 con los valores del factor de bloqueo, densidad ideal y densidad real correspondientes al sistema de ficheros de organización base serial no consecutiva. Adjuntar el desarrollo de los cálculos y explicar de forma razonada los resultados.

Tabla 6. Espacio de almacenamiento con organización serial no consecutiva

Factor de bloqueo	
Densidad ideal	
Densidad real	

3.3 Gestión del Tiempo de Acceso

En este apartado se detallará, de forma teórica, el número mínimo, medio y máximo de accesos a memoria necesarios para los dos procesos de consulta descritos según el sistema de ficheros con organización serial no consecutiva. Se justificarán los valores reflejados en la tabla para cada uno de los casos.

Tabla 7. Tiempos de acceso con organización serial no consecutiva

Clave de consulta	Caso		
	Mejor	Medio	Peor
Seud			
Ape1			
Titu			
Listado			

3.4 Conclusiones de la Organización Serial No Consecutiva

Se ofrecerá una comparativa entre los resultados obtenidos en la Gestión del Espacio de Almacenamiento y la Gestión del Tiempo de Acceso de esta organización base y de la organización original, describiéndose de forma justificada los aspectos en los que la nueva organización mejora a la anterior y aquellos en las que empeora (si se da el caso).

4. **Sistema con Organización Direccionada**

El principal objetivo de esta versión es mejorar el acceso a la información del sistema de ficheros. Para ello, el fichero se estructurará de acuerdo a una organización direccionada tomando como clave de direccionamiento el campo '*seudónimo*'. Se diseñarán los cubos en que se estructura el fichero, se elegirá la función de transformación que se aplicará para dispersar los registros en el fichero y se definirá una estrategia para gestionar los desbordamientos.

4.1 **Diseño del Cubo**

Indicar el espacio de cubo elegido y justificar su elección.

Se indicará (si procede) cuáles son los elementos de datos (directorios, marcas, etc.) agregados al cubo para hacer posible su control (marca de fin de fichero, directorios, etc.). Se especificará en qué lugar del cubo aparece cada uno de ellos (inicio del fichero, fin del fichero, etc.), cuál es su tamaño, tipo, codificación, etc. Se describirá, además, la función de cada uno de ellos en el cubo.

Tabla 8. Elementos de datos de control del cubo

Elemento de datos	Tipo	Codificación / Estructura	Frecuencia de repetición	Tamaño (bytes)

Indicar el tamaño de cubo, adjuntando los cálculos necesarios.

4.2 **Dimensionado del Fichero**

El objetivo es superar el 50% de ocupación del fichero. Se calculará, de forma justificada, el número de cubos de que debe constar el fichero. El estudio se realizará sobre los 1000 registros que se aportan para la realización de la práctica.

4.3 **Elección de la Función de Transformación**

Se compararán el comportamiento de las funciones de transformación residuo y truncamiento a la hora de dispersar los registros del fichero. Se explicarán de forma justificada los resultados obtenidos y se elegirá de forma razonada una de ellas. El estudio se realizará sobre los 1000 registros que se aportan para la realización de la práctica.

Tabla 9. Estudio de la dispersión

	Residuo	Truncamiento
Registros desbordados (%)		
Cubos vacíos (%)		
Ocupación media (%)		

4.4 **Gestión de Desbordamientos**

La versión básica del sistema ficheros deberá implementar gestión de desbordamientos por área de desbordamientos serial. Cualquier otra estrategia será considerada una mejora sobre esta versión básica. Se explicará la estrategia implementada, los elementos de datos

adicionales que requiere, así como las repercusiones que tiene en cuanto a espacio de almacenamiento y tiempo de acceso.

4.5 Gestión del Espacio de Almacenamiento

Rellenar la Tabla 10 con los valores de densidad ideal y densidad real correspondientes al sistema de ficheros de organización base direccionada. Adjuntar el desarrollo de los cálculos y explicar de forma razonada los resultados.

Tabla 10. Espacio de almacenamiento con organización direccionada

Densidad ideal	
Densidad real	

4.6 Gestión del Tiempo de Acceso

En este apartado se detallará, de forma teórica, el número mínimo, medio y máximo de accesos a memoria necesarios para los dos procesos de consulta descritos según el sistema de ficheros con organización direccionada. Se justificarán los valores reflejados en la tabla para cada uno de los casos.

Tabla 11. Tiempos de acceso con organización direccionada

Clave de consulta	Caso		
	Mejor	Medio	Peor
Seud			
Ape1			
Titu			
Listado			

4.7 Comparación con las Organizaciones Anteriores

Se ofrecerá una comparativa entre los resultados obtenidos en la Gestión del Espacio de Almacenamiento y la Gestión del Tiempo de Acceso de esta organización base y de las organizaciones base anteriormente tratadas, describiéndose de forma justificada los aspectos en los que la nueva organización mejora a las anteriores y aquellos en las que empeora (si se da el caso).

5. Sistema con Organización Auxiliar: Índices

Se añadirán al sistema de ficheros anterior índices para el título de episodios y primer apellido del personaje. Los índices serán, al menos, seriales y exhaustivos. Se justificará la estructura de los índices elegida y, de identificarse alguna estructura mejor que no hubiese sido implementada por razones de costes de desarrollo, será también descrita.

Tabla 12. Índices implementados

Campo	Nombre Del Índice	Tipo (*)	Tamaño KB
Titu			
Ape1			

(*) Carácter (primario/secundario); Densidad (denso/no denso); Orden (ordenado/desordenado); Organización (serial, listas, arbórea,...); Recubrimiento (exhaustivo/parcial); Invertido (Si/No); ...

5.1 Gestión del Espacio de Almacenamiento

Rellenar la Tabla 13 con los valores del factor de bloqueo, densidad ideal y densidad real correspondientes al sistema de ficheros de organización auxiliar. Adjuntar el desarrollo de los cálculos y explicar de forma razonada los resultados.

Tabla 13. Espacio de almacenamiento con organización auxiliar indexada

Densidad ideal	
Densidad real	

5.2 Gestión del Tiempo de Acceso

En este apartado se detallará, de forma teórica, el número mínimo, medio y máximo de accesos a memoria necesarios para los tres procesos de consulta descritos según el sistema de ficheros con organización auxiliar indexada. Se justificarán los valores reflejados en la tabla para cada uno de los casos.

Tabla 14. Tiempos de acceso con organización auxiliar indexada

Clave de consulta	Caso		
	Mejor	Medio	Peor
Seud			
Ape1			
Titu			
Listado			

5.3 Comparación con las Organizaciones Anteriores

Se ofrecerá una comparativa entre los resultados obtenidos en la Gestión del Espacio de Almacenamiento y la Gestión del Tiempo de Acceso de esta organización base y de las organizaciones base anteriormente tratadas, describiéndose de forma justificada los aspectos en los que la nueva organización mejora a las anteriores y aquellos en las que empeora (si se da el caso).

6. **Conclusiones**

Realizar un resumen del trabajo realizado. Se describirá, en términos de gestión de espacio y de tiempo de acceso, las ventajas y virtudes de cada una de las alternativas implementadas. Se describirán posibles mejoras futuras que pudieran realizarse sobre el sistema implementado y qué implicarían.

7. Anexo 1: Ejemplo de Diseño Lógico de un Registro

Se supondrá que quiere ser representado el diseño lógico de un registro con las siguientes características:

- El registro consta de dos campos, CAMPO1 y CAMPO2. El primero es una cadena de caracteres de longitud variable menor de 1000 caracteres y frecuencia de reiteración indefinida (aunque menor de 256).
- CAMPO2, por su parte, consta de dos campos: CAMPO2_1 y CAMPO2_2.
- CAMPO2_1 es una cadena de hasta 400 caracteres, que sólo puede tomar 5 posibles valores. Este campo es reiterativo (pero siempre se repite menos de 256 veces).
- CAMPO2_2 es una cadena de longitud fija de 15 caracteres.

El diseño lógico de este registro es representado como sigue:

```
Registro{
    CAMPO1 (1000)*,
    CAMPO2{
        CAMPO2_1(400)*,
        [CAMPO2_2(15)]
    }
}
```

8. Anexo 2: Ejemplo de Representación del Diseño Físico-Lógico de un Registro

Se supondrá el registro lógico del apartado Anexo 1: Ejemplo de Diseño Lógico de un Registro (más atrás). En este caso, el registro será diseñado de la siguiente forma:

- CAMPO1 se representará como un grupo repetitivo que incluirá una marca de repetición a su inicio de un byte. Cada elemento del grupo repetitivo será un vector que constará de una marca de longitud del campo de 2 bytes y de su valor: una cadena de caracteres de longitud variable.
- CAMPO2 será representado como un vector que contendrá CAMPO2_1 y CAMPO2_2.
- CAMPO2_1 será tratado como un grupo repetitivo que incluirá a su inicio una marca de repetición de un byte. Cada elemento del grupo repetitivo será un valor enumerado con un byte.
- CAMPO2_2 será representado como un vector que incluye una marca de existencia binaria (representada con un byte) y su valor (cuando existe): una cadena de 15 caracteres de longitud fija.

Este diseño es representado según la siguiente tabla:

Elemento de datos	Tipo	Codificación / Estructura	Frecuencia de repetición	Tamaño (bytes)
GRCAMPO1	Grupo repetitivo	#VCAMPO1 VCAMPO1	1	0
#VCAMPO1	Marca de reiteración	Numérico	1	1
VCAMPO1	Vector	LCAMPO1 CAMPO1	#VCAMPO1	0
LCAMPO1	Marca de longitud	Numérico	1	2
CAMPO1	Campo	Alfanumérico	1	LCAMPO1
VCAMPO2	Vector	GRCAMPO2_1 VCAMPO2_2	1	0
GRCAMPO2_1	Grupo repetitivo	#CAMPO2_1 CAMPO2_1	1	0
#CAMPO2_1	Marca de reiteración	Numérico	1	1
CAMPO2_1	Campo	Enumerado	#CAMPO2_1	1
VCAMPO2_2	Vector	ECAMPO2_2 CAMPO2_2	1	0
ECAMPO2_2	Marca de existencia	Binario	1	1
CAMPO2_2	Campo	Alfanumérico	1	15