

Estructuras de Archivos



- **Archivo**

- Es una colección de registros lógicamente relacionados.

- **Registro**

- Es una estructura de campos o elementos de información lógicamente relacionados.

Clasificación de Archivos

Trabajo

- Es un archivo temporal

Texto

- Cualquier archivo de usuario

Reporte

- Visualización de datos de forma temporal

Configuración

- Permite modificar la información

Transacción

- Importante para base de datos

Programa

- Cualquier archivo escrito en un lenguaje de programación

Diferentes tipos de registros

- Entidad
- Atributo
- Datos
- Índice

Índice

- Los registros en el archivo indexado pueden estar a su vez almacenados siguiendo un orden.
- Un archivo puede tener varios índices según diferentes claves de búsqueda. Si el archivo que contiene los registros está ordenado secuencialmente, el índice cuya clave de búsqueda especifica el orden secuencial del archivo es el índice con agrupación.
- Los índices con agrupación también se llaman índices primarios.
- Los índices cuyas claves de búsqueda especifican un orden diferente del orden secuencial del archivo se llaman índices sin agrupación o secundarios.

Estructura
de Datos

Agrupación

→ Primarios

Sin agrupación

→ Secundarios

Orden secuencial

→ Apuntador siguiente

Índice Denso

- Únicamente al primer registro que coincida con el nombre.
- Aparece un registro índice por cada valor de la clave de búsqueda en el archivo.
- En un índice denso con agrupación, el registro índice contiene el valor de la clave de un apuntador al primer registro con ese valor de la clave de búsqueda.
- El resto de registros con el mismo valor de la clave de búsqueda se almacenan consecutivamente después del primer registro, dado que ya que el índice es con agrupación, los índices se ordenan sobre la misma clave de búsqueda.
- La implementación de índices densos pueden almacenar una lista de apuntadores a todos los registro con el mismo valor de la calve de búsqueda, esta no es esencial para los índices con agrupación.

Inserción de un índice denso

1. Se realiza una búsqueda usando el valor de la clave de búsqueda del registro a insertar.
2. Si el valor de la clave de búsqueda no aparece en el índice, el sistema inserta en este un registro índice con el valor de la clave de búsqueda en la posición adecuada
3. En caso contrario se ejecutan la siguientes acciones:
 1. Si el registro índice almacena apuntadores a todos los registros con el mismo valor de la clave de búsqueda, el sistema añade un apuntador al nuevo registro en el índice.
 2. En caso contrario , el registro índice almacena un apuntador solo hacia el primer registro con el valor de la clave de búsqueda. El sistema sitúa el registro insertado después de los otros con los mismos valores de la clave de búsqueda.

Eliminación de un índice denso

1. Para borrar un registro, primero se busca el índice a borrar. Si el registro borrado era el único registro con ese valor de la clave de búsqueda, el sistema borra el registro índice correspondiente del índice.
2. En caso contrario, se ejecutan las siguientes acciones:
 1. Si el registro índice almacena apuntadores a todos los registros con el mismo valor de la clave de búsqueda, el sistema borra el registro.
 2. En caso contrario; el registro índice almacena un apuntador solo al primer registro con el valor de la clave de búsqueda. En este caso, si el registro borrado era el primer registro con el valor de la clave de búsqueda, el sistema actualiza el registro índice para apuntar al siguiente registro.

Diccionario de Datos

- Es donde se almacenan las descripciones de los datos y sea accesible para los usuarios.
- Este catalogo es lo que se denomina “diccionario de datos” y contiene información que describe los datos de la base de datos (metadatos).
- Normalmente un diccionario de datos describe entre otros casos:
 - Nombre, tipo y tamaño de los datos (Entidades y Atributos).
 - Relaciones entre datos.
 - Restricciones de integridad sobre los datos.
 - Usuarios autorizados a acceder a los objetos de base de datos.
 - Estadísticas de utilización, tales como la frecuencia de las transacciones y el número de accesos realizados a los objetos de la base de datos.

Elementos de un diseño conceptual

- Entidad
- Atributo
- Relación
- Regla empresarial o de negocios
- Datos de Intersección

Entidad

- Es una persona, lugar, cosa, suceso o concepto sobre el que se recopilan datos. En otras palabras, las entidades son los objetos reales que nos interesan lo suficiente como para capturar y guardar sus datos en una base de datos. Una entidad se representa como un rectángulo en el diagrama. Cualquier cosa que se designe con un nombre puede ser una entidad.

Atributo

- Es un hecho unitario que caracteriza o describe de alguna manera a una entidad.
- Se dice que los atributos son un hecho aislado porque deben ser indivisibles, lo que significa que no pueden dividirse en unidades mas pequeñas que tengan algún significado.

Relaciones

- Son las asociaciones entre las entidades. Como las bases de datos se concentran en guardar datos relaciones, las relaciones se vuelven el pegamento que mantiene unida la base de datos.

Regla Empresarial o de negocios

- Es una política, procedimiento o norma adoptada por una organización. Las reglas de negocios son muy importantes en el diseño de una base de datos porque determinan los controles de que deben aplicarse a los datos.
- Las reglas de negocios implementan en la base de datos como restricciones, que son las reglas formalmente definidas que limitan.

Metadatos de una Entidad

- **Nombre**
 - Identificador propia del objeto que se requiere representar
- **Dirección de la entidad**
 - Representa la dirección de archivo en donde se almacena la entidad
- **Dirección de los atributos**
 - Representa la dirección de archivo en donde se encuentra el primer atributo.
- **Dirección de los datos**
 - Representa la dirección de archivo en donde se encuentra el primer registro de datos.
- **Dirección de la siguiente entidad**
 - Representa la dirección de la siguiente entidad

Nombre	DE	DA	DD	DSIG
--------	----	----	----	------

Metadatos de un Atributo

- **Nombre**
 - Identificador propio del atributo
- **Tipo de dato**
 - Se utilizara Entero (E) y Carácter (C)
- **Longitud de tipo de dato**
 - Representa la longitud en bytes del tipo de dato
- **Dirección del atributo**
 - Representa la dirección de archivo en donde se almacena el atributo

- **Tipo de Índice**

- Identifica el tipo de estructura de datos que se utiliza para representar el índice
 - 0 → sin tipo de índice
 - 1 → clave de búsqueda
 - 2 → índice primario
 - 3 → índice secundario

- **Dirección de índice**

- Representa la dirección de archivo en donde se comienza a almacenar la estructura de datos utilizada para representar el índice

- **Dirección del siguiente atributo**

- Representa la dirección del siguiente atributo que define a la misma entidad.

Nombre	Tipo	Longitud	DA	TI	DI	DSIG
--------	------	----------	----	----	----	------

Registros

Dirección	Atributos	Dirección Siguiente
-----------	-----------	---------------------

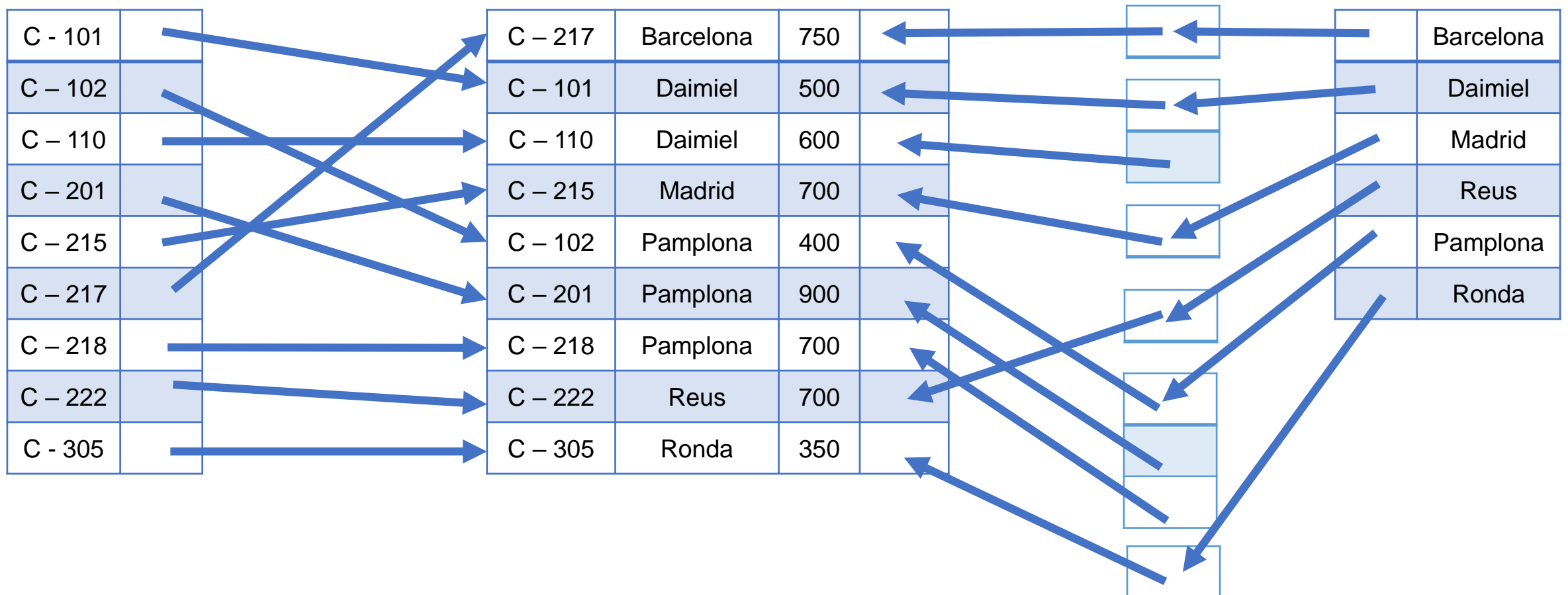
- El tamaño de cada registro se calcula sumando la longitud de todos los atributos que tiene la entidad mas el espacio de su dirección y su apuntador al siguiente

Organización Secuencial Indexado

- Se utiliza este tipo de organización de archivo cuando existe la necesidad tanto de acceder a los registros secuencialmente por algún valor de la clave de búsqueda (llave primaria o secundaria), como de accederlos individualmente.
- Un archivo secuencial indexado **proporciona la combinación de tipos de acceso que manejan un archivo secuencial** y un archivo relativo o de acceso directo.

Estructura Lógica de un archivo secuencial indexado

- En este tipo de organización de archivos se dispone de una tabla en la que aparecen ordenados los valores de la clave de búsqueda y asociada a cada uno de ellos la dirección del registro correspondiente.



Diferentes archivos

- Registro de Datos
 - Archivo de Datos
 - Alumno.dat
- Registro de Índices
 - Alumno.idx

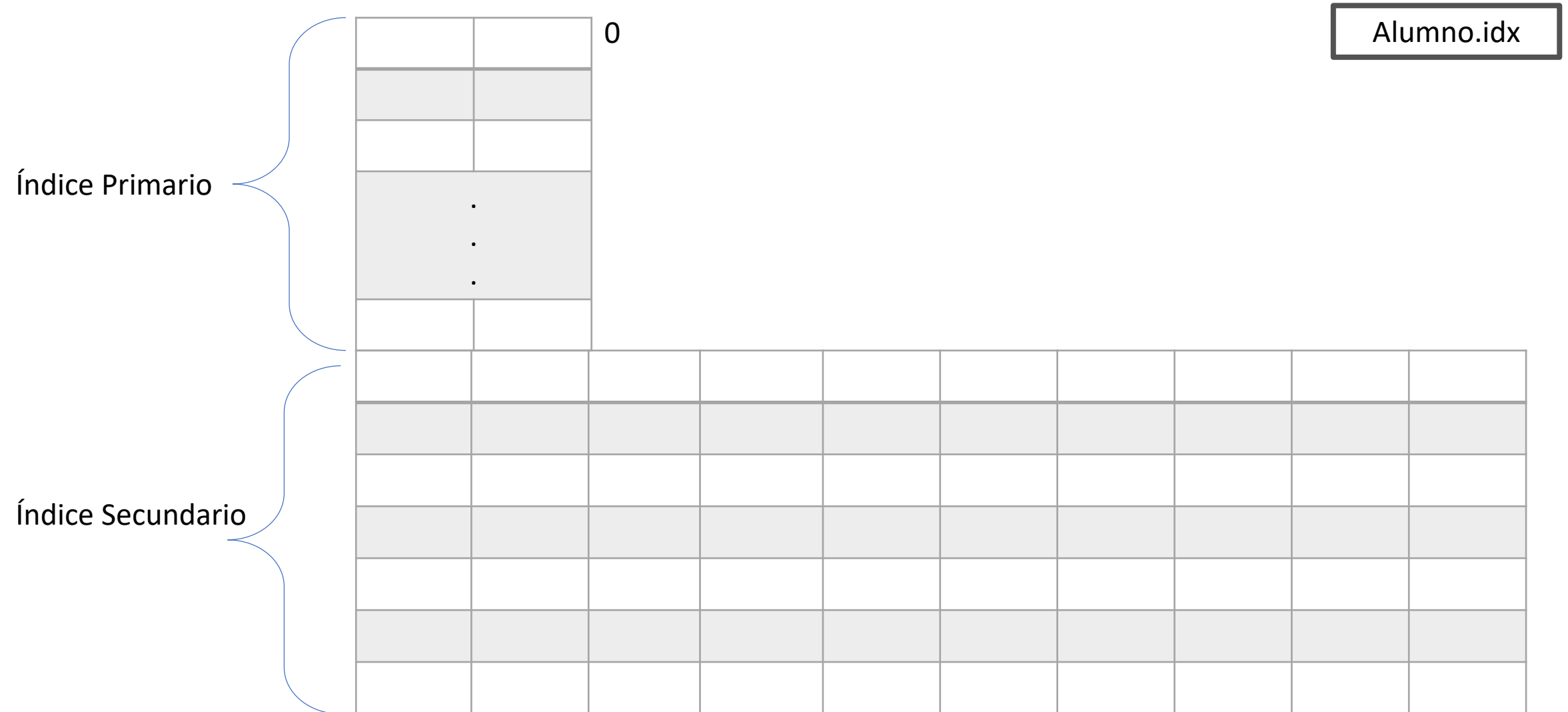
cve_única	nombre	cve_carrera	genero
830022	JOSE	15	M
292122	JORGE	14	M
993323	ABRAHAM	15	M
392367	MARIA	14	F
199221	JOSUE	15	M
491235	ALICIA	14	F
549323	EDUARDO	14	M
735232	GERARDO	15	M
319222	ISABEL	14	F
229906	SANDRA	15	F
669240	FRANCISCO	15	M
673776	MIGUEL	15	M

Entidad: Alumno			
Atributos	TD	LONG	INDICE
cve_única	E	4	2
nombre	C	15	1
cve_carrera	E	4	3
genero	C	1	0

Diccionario de Datos

8 Alumno	8	70	-1	-1		
cve_única	E	4	70	2	-1	133
nombre	C	15	133	1	-1	196
cve_carrera	E	4	196	3	-1	259
genero	C	1	259	0	-1	-1

Registros (Alumno.dat y Alumno.idx)



REGISTRO DE DATOS

8	Alumno	8	70	80	-1
---	--------	---	----	----	----

Alumno.dat

[illegible]

Alumno.idx

292122	40
830022	0
993323	80

14	40			
15	0	80		

Archivos de índices de árbol B+

- La estructura de índice de árbol B+ es la más entendida de las estructuras de índices que mantienen su eficiencia a pesar de la inserción y borrado de datos. Un índice de árbol B+ toma la forma de un árbol equilibrado donde los caminos de la raíz a cada hoja del árbol son de la misma longitud.
- Cada nodo que no sea hoja tiene entre $n/2$ y n hijos, donde n es fijo para cada árbol.

Estructura de un árbol B+

P1	K1	P2	K2	...	Pn-1	Kn-1	Pn
----	----	----	----	-----	------	------	----

Nodo de un árbol B+

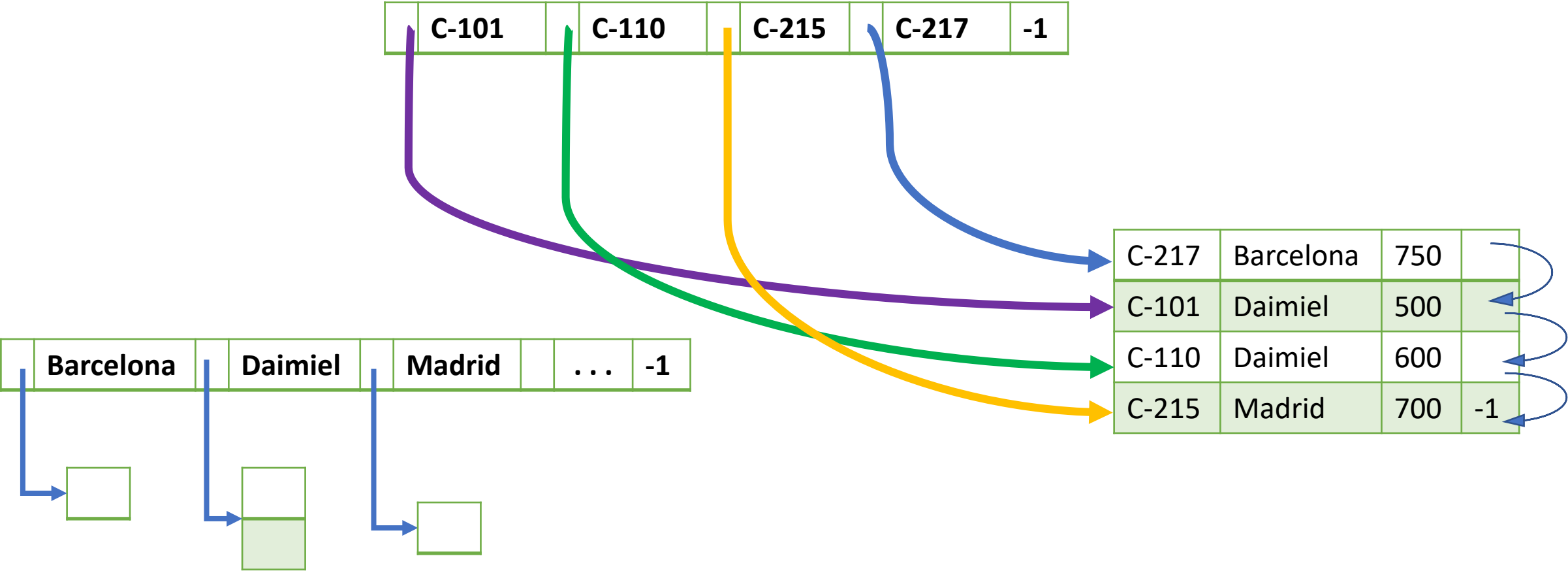
- En la figura se muestra un nodo típico de un árbol. Puede contener hasta $n-1$ claves de búsqueda (k_1, k_2, \dots, k_{n-1}) y n apuntadores (P_1, P_2, \dots, P_n). Los valores de las claves de búsqueda se mantienen ordenados; así, si $i < j$, entonces $k_i < k_j$.
- Considere primero la estructura de los nodos hoja.
- Para $i = 1, 2, \dots, n-1$, el apuntador P_i apunta, o bien a un registro del archivo con valor de clave de búsqueda k_i , o bien a una lista de apuntadores, cada uno de los cuales apunta a un registro del archivo con valor de la clave de búsqueda k_i .

Nodo hoja (Pk)

	C-101	C-110	C-215	C-217	-1
--	-------	-------	-------	-------	----

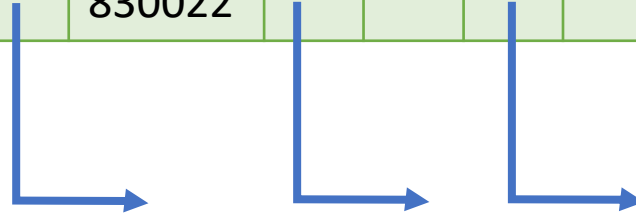
	Barcelona	Daimiel	Madrid	...	-1
--	-----------	---------	--------	-----	----

C-217	Barcelona	750	
C-101	Daimiel	500	
C-110	Daimiel	600	
C-215	Madrid	700	-1



Atributo

TN	DH	P1	K1	P2	K2	P3	K3	P4
H			830022					-1



- TN → Tipo de Nodo
- DN → Dirección de Nodo
- K → Clave de búsqueda
- P → Apuntador

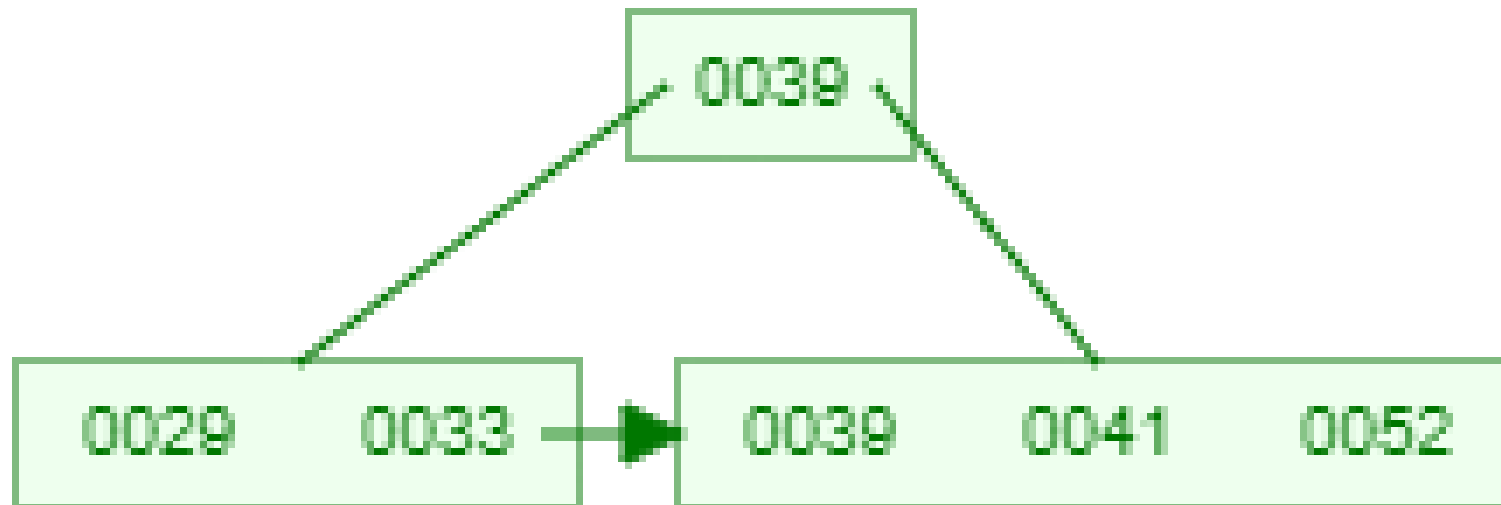
Tipos de Nodo

- Hoja
- Intermedio
- Raíz

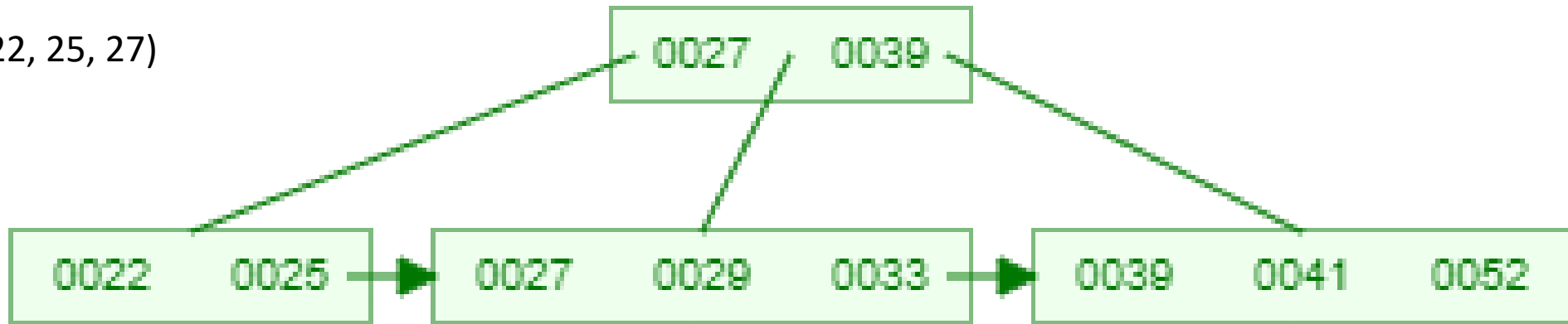
Insertar (29,33,39,41)



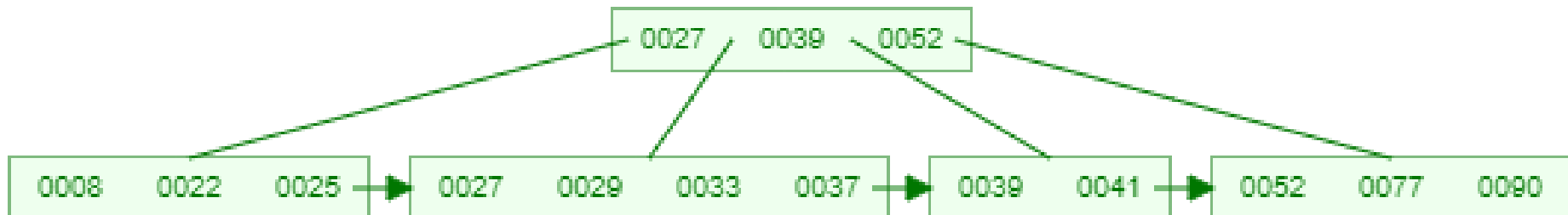
Insertar (52)



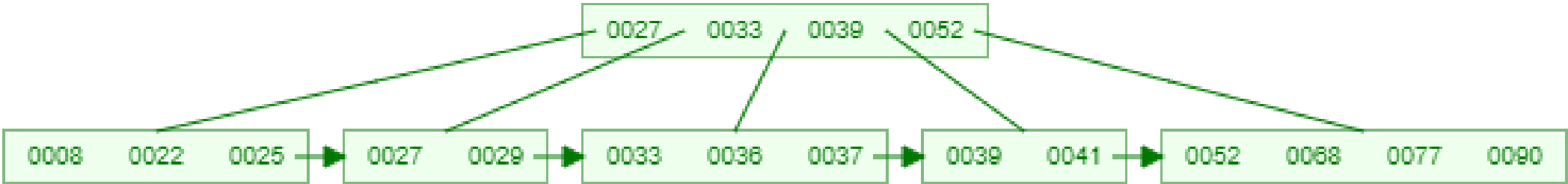
Insertar (22, 25, 27)



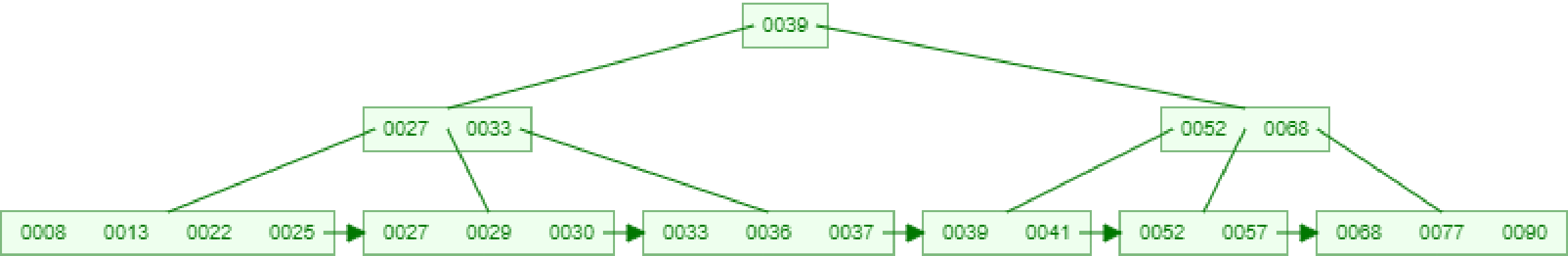
Insertar (37, 90, 8, 77)



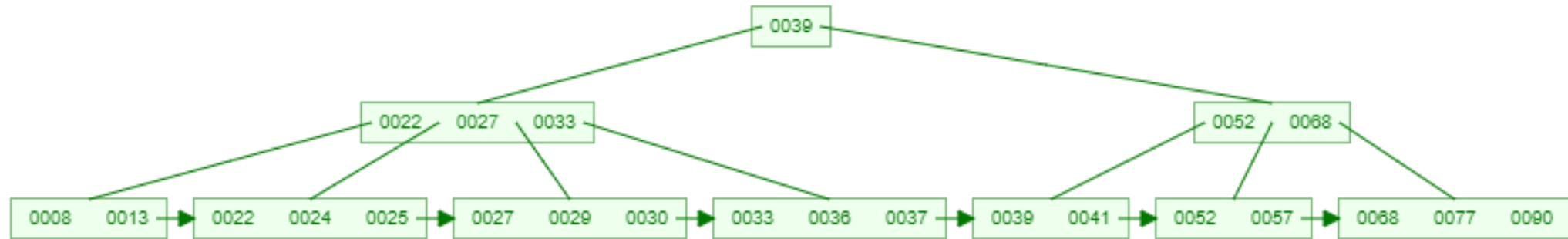
Insertar (68, 36)



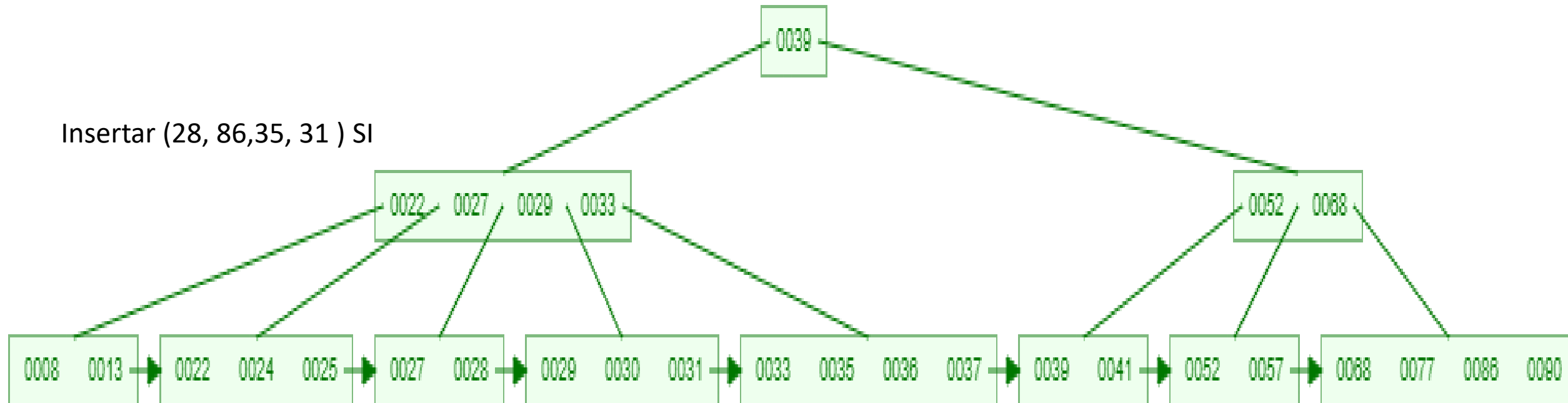
Insertar (30, 13, 57) SI



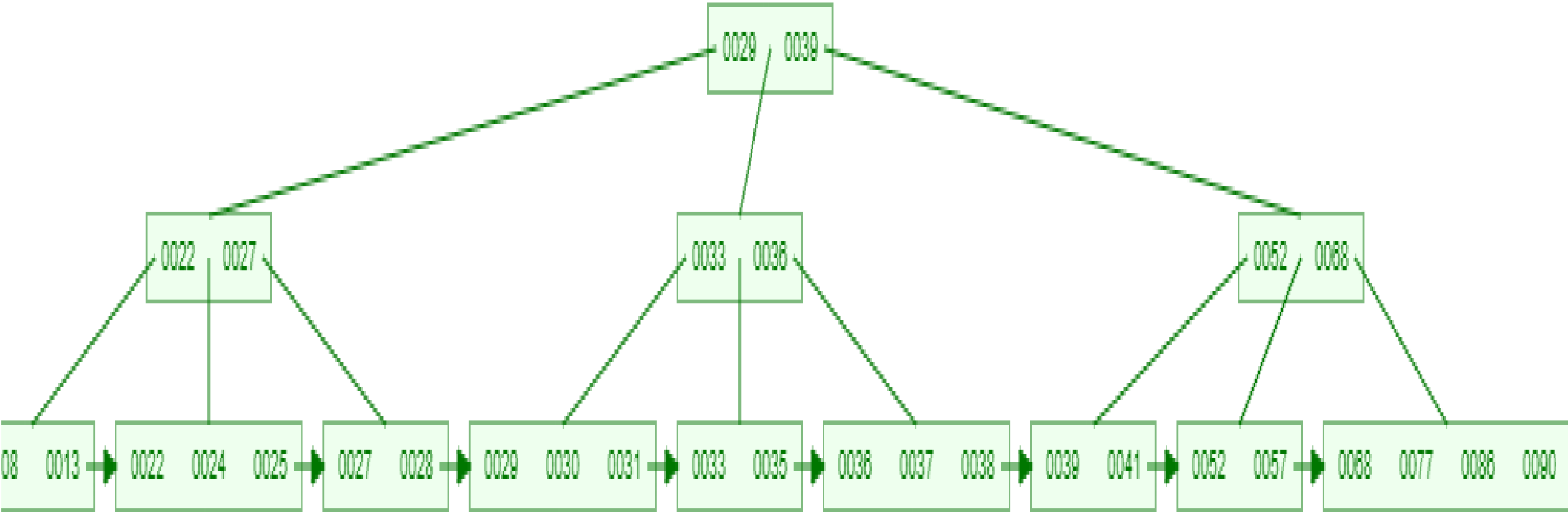
Insertar (24) SI



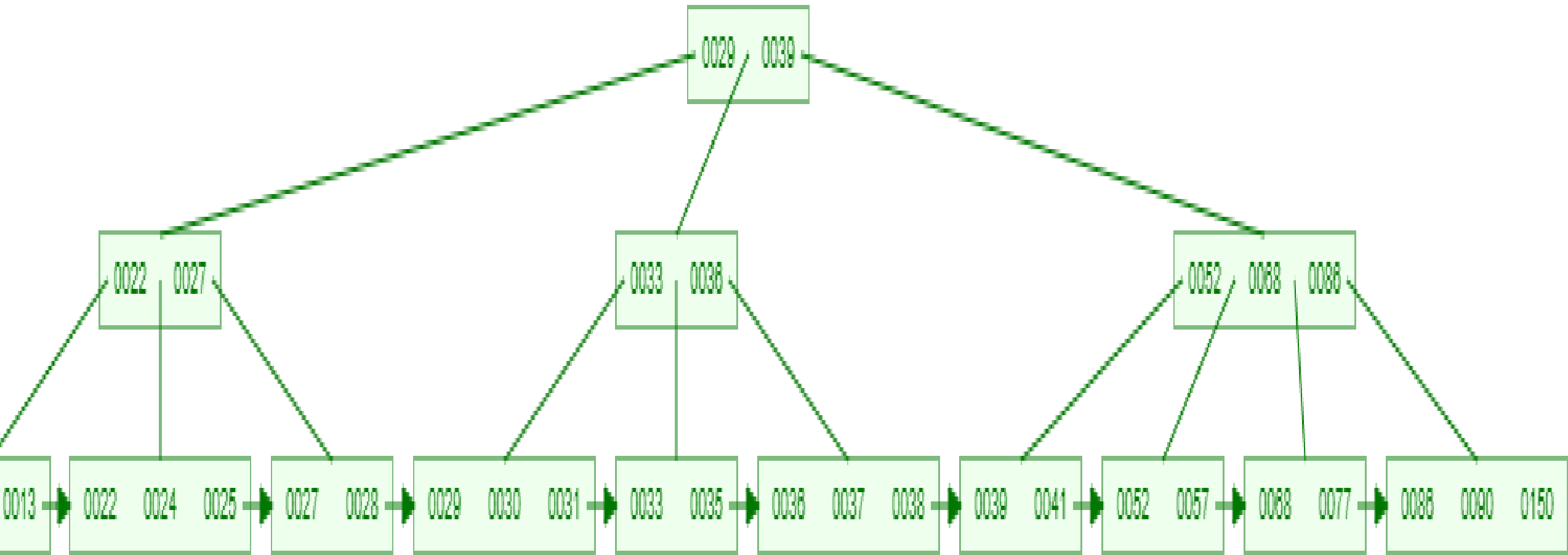
Insertar (28, 86,35, 31) SI



Insertar (38)



Insertar (150)



Actualización en arboles B+

- El borrado y la inserción son operaciones más complicadas que las búsquedas, ya que puede ser necesario dividir un nodo como resultado de una inserción, o fusionar dos nodos si uno se volviera demasiado pequeño.
- Cuando se divide o fusiona un par de nodos, se debe asegurar el equilibrio del árbol.
- Inserción:
 - Se busca un nodo hoja donde tendría que aparecer el valor de la clave de búsqueda.
 - Si el valor de la clave de búsqueda ya aparece en el nodo hoja, se inserta un nuevo registro en el archivo y, si es necesario un apuntador en la lista, en caso de ser índice secundario.

-
- Borrado:
 - Se busca el registro a borrar y se elimina del archivo. Si no existe una lista asociada con el valor de la clave de búsqueda o si la lista se queda vacía como resultado del borrado, se borra el valor de la clave búsqueda del nodo hoja.
-
- * El apuntador que está en el tipo de atributo debe de apuntar al primer nodo del árbol.
 - * Las claves de búsqueda se insertan en orden.

-
- B) Modificación de datos:
 - Esta acción consiste en buscar un registro específico y cambiar el contenido de uno o de varios campos del registro. En el caso de las modificaciones, se hace el recorrido por los índices hasta llegar al registro que se desea modificar.
 - Una vez que se encontró el registro se podrá cambiar el contenido de todos los campos.
 - C) Eliminación de registros:
 - Esta acción consiste en eliminar un registro en específico, eliminando tanto del archivo de datos como de los índices.

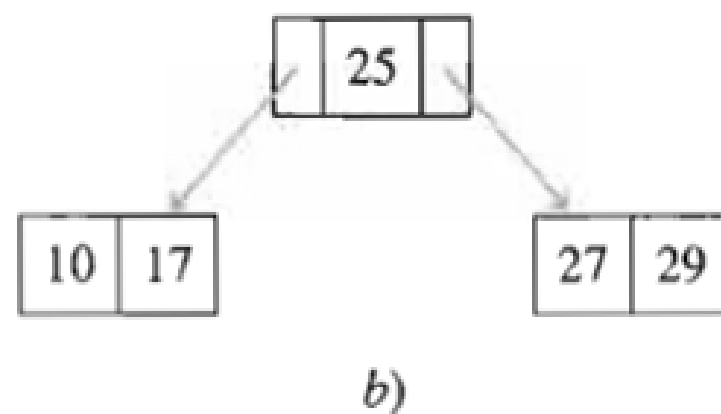
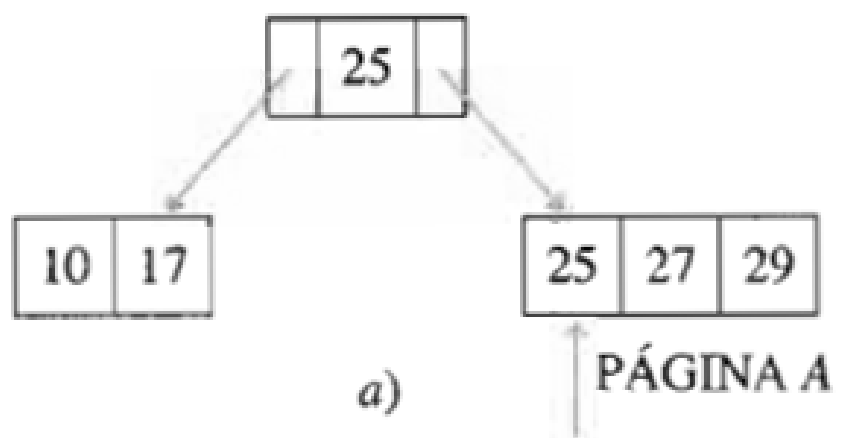
• Cve_única	nombre	cve_carrera
• 413	JUAN	74
• 721	JORGE	75
• 983	SUSANA	63
• 243	MONICA	74
• 332	JUAN	72
• 108	GRABRIELA	75
• 536	JORGE	74
• 295	RUBEN	63
• 133	ROSARIO	63

Eliminación en Árboles B+

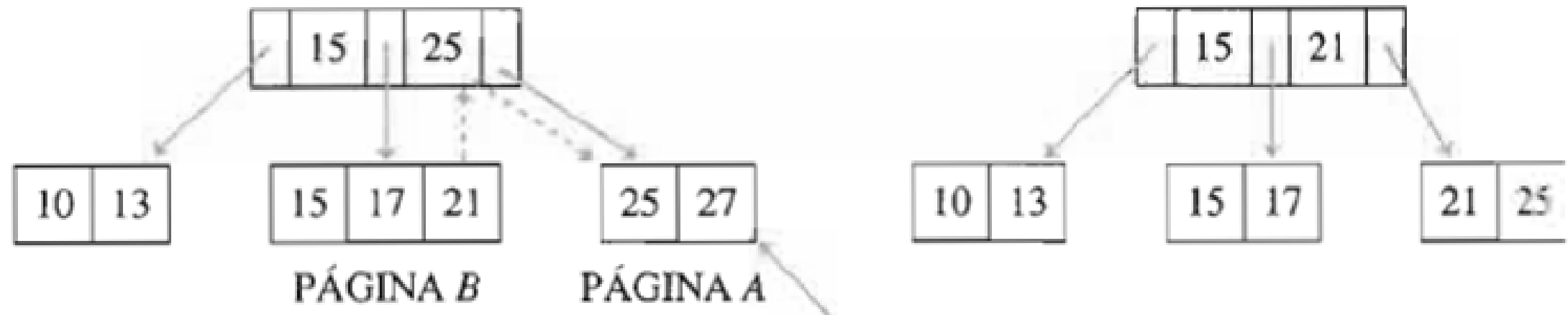
- La eliminación de árboles B+ es más simple que la operación de borrado en árboles B. Esto ocurre porque las claves que se deben eliminar siempre se encuentran en las páginas de la hoja. En general se deben distinguir estos casos:
 - **CASO 1**
 - Si al eliminar una clave m queda mayor o igual a d , entonces termina la operación de borrado. Las claves de las páginas raíz o internas no se modifican por más que sean una copia de la clave eliminada en las hojas.
 - **CASO 2**
 - Si al eliminar una clave m queda menor a d , entonces se debe realizar una redistribución de claves, tanto en el índice como en las páginas hojas. Cuando se cambia la estructura del árbol, se quitan aquellas claves que quedaron en los nodos interiores luego de haber eliminado su correspondiente información en los nodos hoja.

-
- Puede suceder que al eliminar una clave y al realizar una redistribución, la altura del árbol disminuya en una unidad.

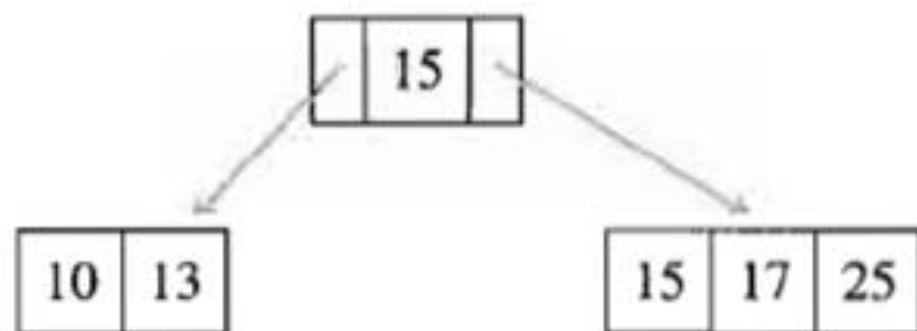
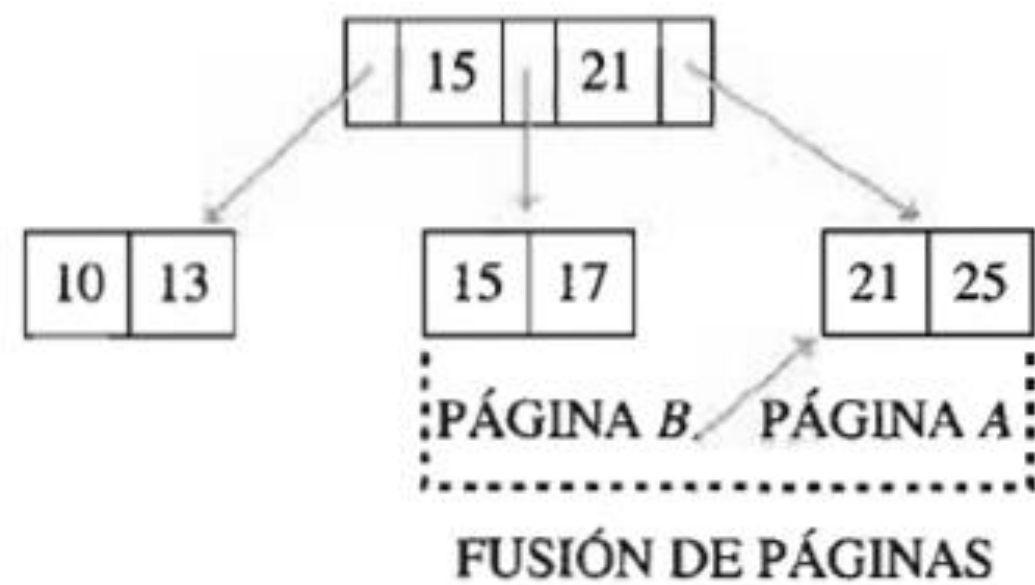
ELIMINACIÓN: CLAVE 25



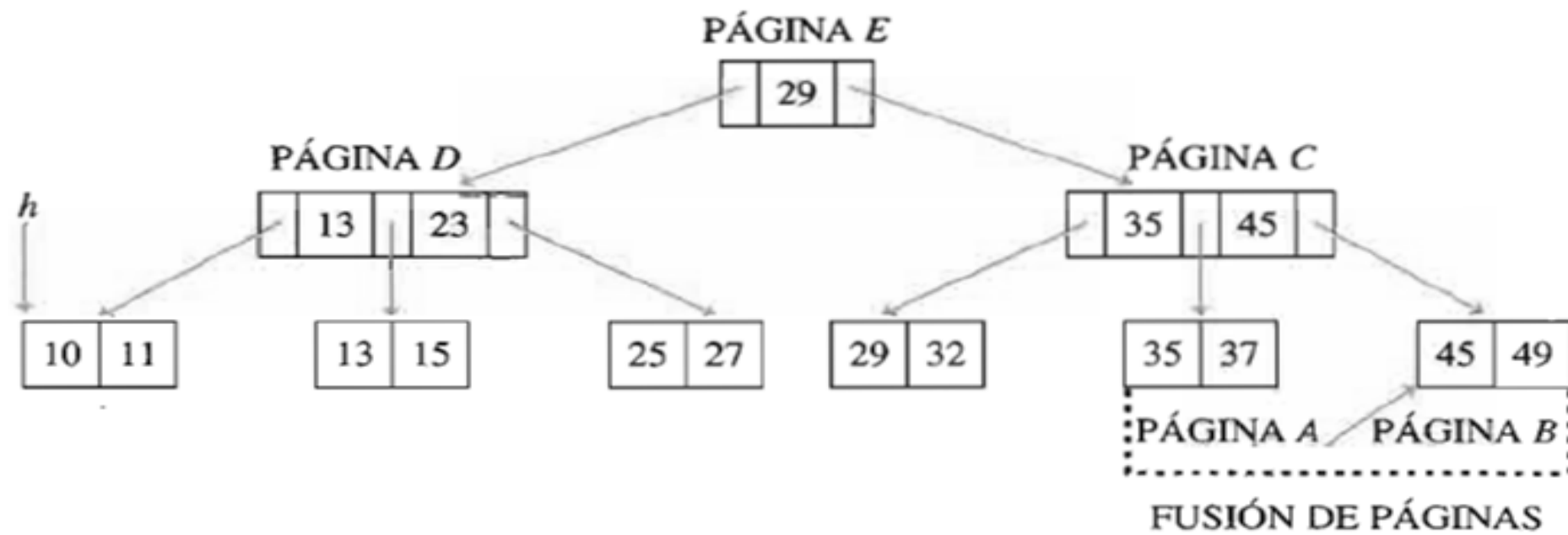
ELIMINACIÓN: CLAVE 27



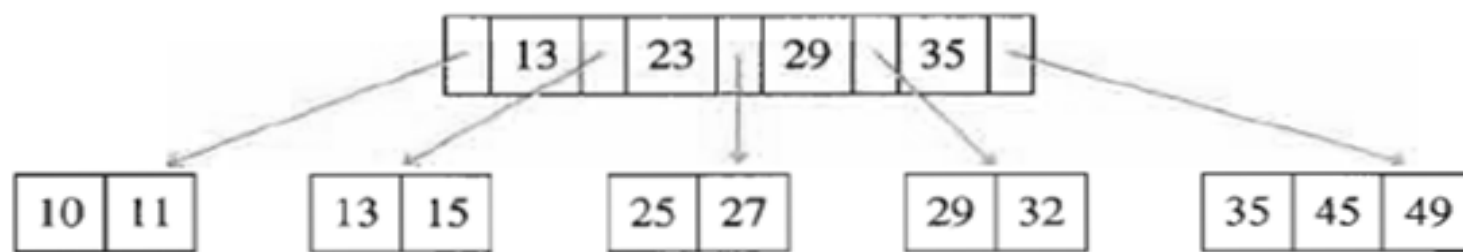
ELIMINACIÓN: CLAVE 21



ELIMINACIÓN: CLAVE 37



a)



b)