ARBOL

ESTRUCTURAS DE DATOS y ALGORITMOS LCC TUPW

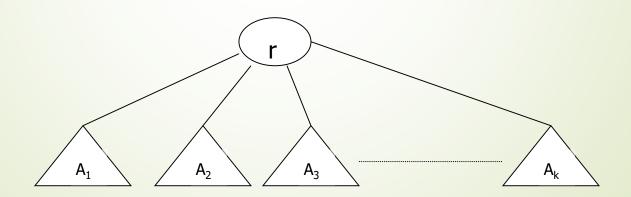
Objetivos

- ➤ Conocer distintos tipos de árboles y sus aplicaciones habituales.
- Realizar un análisis comparativo entre los mismos, a partir de la evaluación de costo de ejecución de los algoritmos que los manipulan.
- Construir los TADs: Árbol BB, AVL, B y Montículo Binario.

ARBOL

Un **ÁRBOL** es una colección o conjunto de nodos. La colección puede:

- 1 Estar vacía
- 2 Si no está vacía, consiste de un nodo distinguido r, conocido como *raíz*, y cero o más (sub)árboles A1, A2,....,Ak, cada uno de los cuales tiene su raíz conectada a r por medio de una *arista* (o *rama*) dirigida.



ARBOL

- La raíz de cada subárbol es un *hijo* o *descendiente directo* de r, y r es el *padre* o *antecesor directo* de cada raíz de los subárboles.
- ➤ Camino de un nodo ni a otro nk: secuencia de nodos n1, n2,, nk , tal que ni es el padre de ni+1 para 1<=i<k. En un árbol existe solamente un camino desde la raíz a cada nodo.
- ➤ Si hay un camino entre los nodos n1 y n2, entonces n1 es *antecesor* de n2 y n2 es *descendiente* de n1.
- Longitud de camino de un nodo ni a otro nk. Número de aristas que forman el camino, o número de nodos menos 1 que forman la secuencia. Existe un camino de longitud cero desde cada nodo a sí mismo.
- Nivel de un nodo: si el nodo x está en el nivel i, entonces sus descendientes directos están en el nivel i+1. La raíz de un árbol, se define como localizada en el nivel 1.
- > Profundidad o Altura del árbol: máximo de los niveles de todos los nodos del árbol.
- > Grado de un nodo: Número de descendientes directos de un nodo.
- > Grado del árbol: Grado máximo en todos los nodos.
- > Nodo hoja: nodo de grado 0.
- Nodo Interior: Nodo no hoja.

Árbol Binario

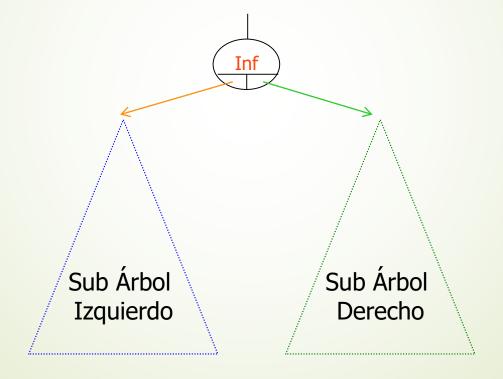
Un árbol ordenado de grado 2 se denomina **Árbol Binario**.

Este árbol se define como un conjunto finito de elementos (nodos) que:

- es vacío o
- consta de una raíz (nodo) con dos árboles binarios disjuntos llamados subárbol izquierdo y derecho de la raíz.

Árbol Binario Representación

Representación enlazada



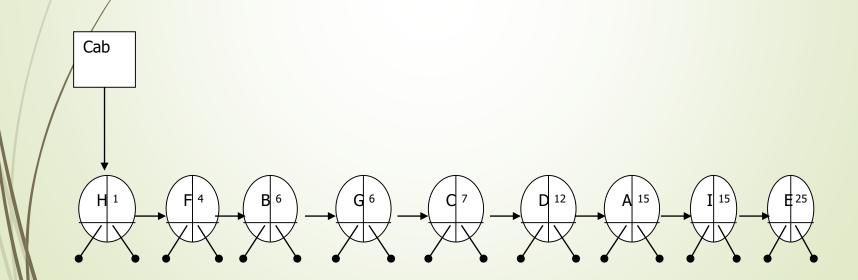
Generación de Códigos de Huffman

Códigos de longitud variable

Caracter	A	В	С	D	E	F	G	Н	I
Frecuencia	15	6	7	12	25	4	6	1	15

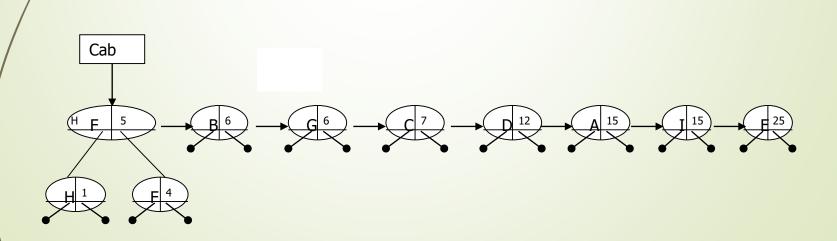
Generación de Códigos de Huffman

Paso 1: Generar una lista de árboles binarios. Cada raiz, inicialmente, contiene un caracter y su frecuencia de ocurrencia. La lista está ordenada en forma creciente por las frecuencias.

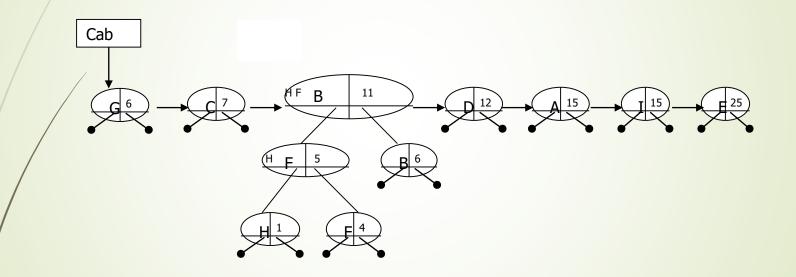


Generación de Códigos de Huffman

- Paso 2: Generar repetidamente árboles binarios a partir de aquellos con frecuencias menores, e insertarlos nuevamente en la lista ordenada. La frecuencia asociada a la raíz de cada nuevo árbol es la *suma* de las frecuencias de sus subárboles.

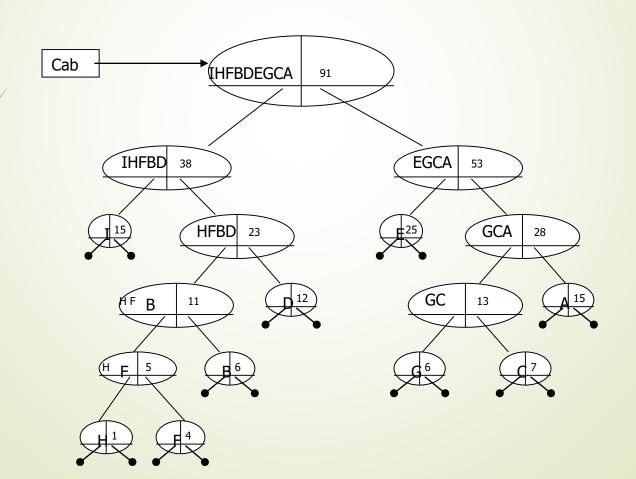


Generación de Códigos de Huffman



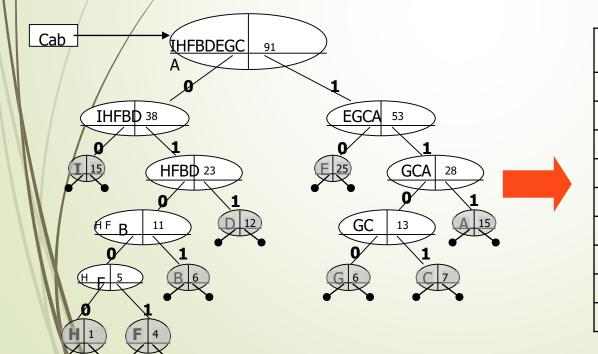
Esta secuencia de supresiones, síntesis e inserciones se realiza hasta el momento en que en la lista queda una sola celda, que corresponde a la raíz del árbol binario que tiene a todos los caracteres del diccionario de entrada como nodos hoja.

Generación de Códigos de Huffman



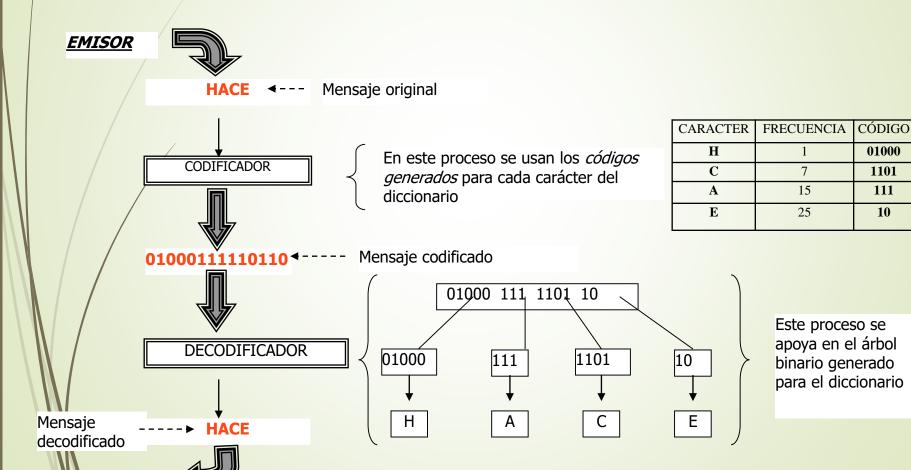
Generación de Códigos de Huffman

Paso 3: Generar el código de longitud variable correspondiente a cada carácter, hoja del árbol, concatenando ceros y unos, según se atraviese en el árbol una rama izquierda o una rama derecha.



CARACTER	FRECUENCIA	CÓDIGO
Н	1	01000
F	4	01001
В	6	0101
G	6	1100
C	7	1101
D	12	011
A	15	111
I	15	00
E	25	10

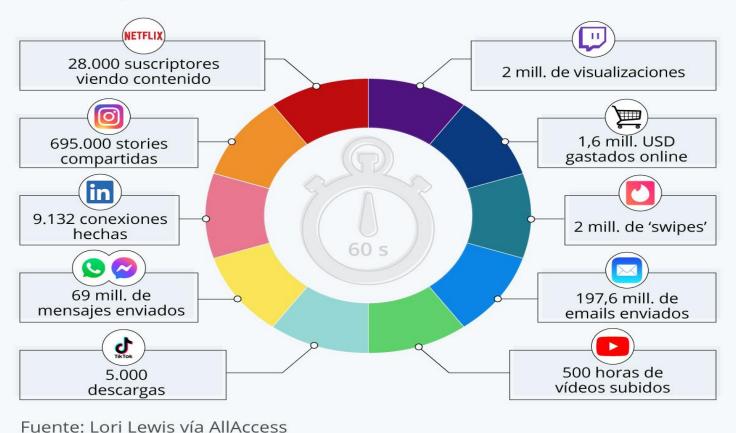
Codificación-Decodificación



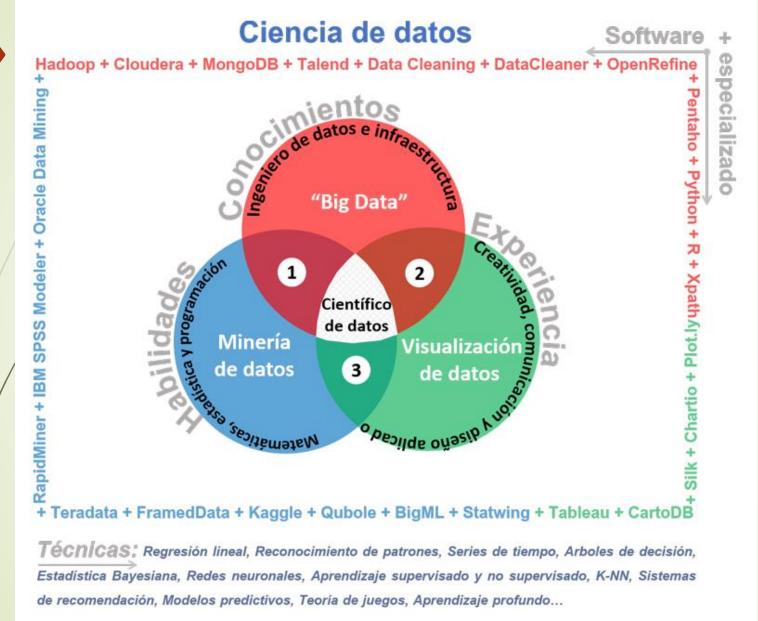
RECEPTOR

Esto sucede en Internet en un minuto

Estimación de una selección de actividades y datos generados online en un minuto en 2021

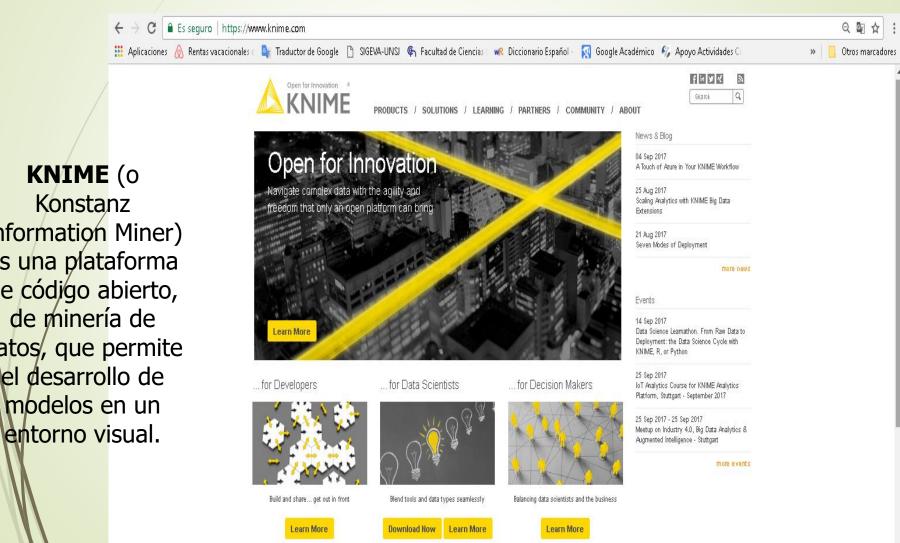


statista 🚄

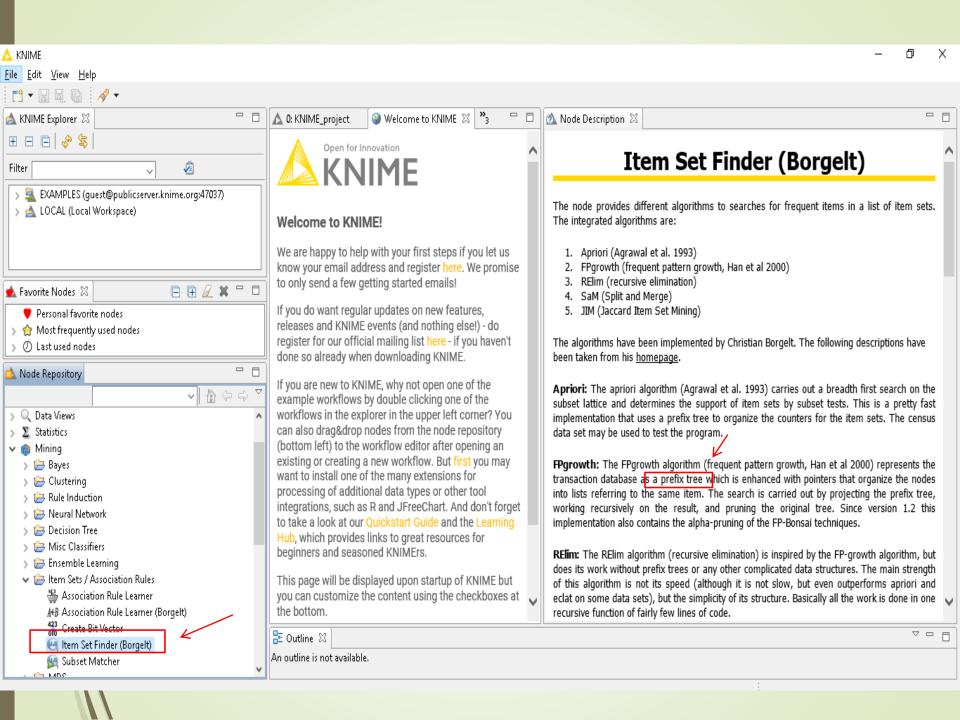


http://www.revista.unam.mx/vol.18/num7/art53/index.html

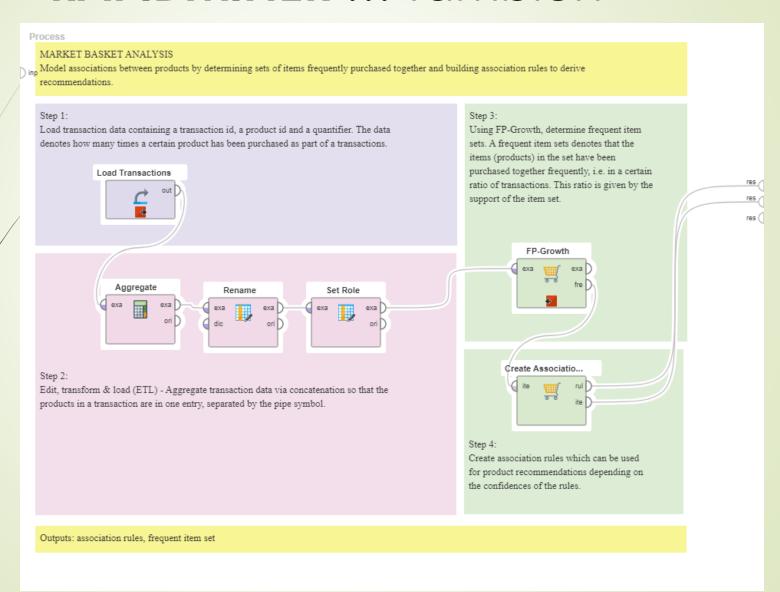
Minería de Datos-KNIME



KNIME (o Konstanz Information Miner) es una plataforma de código abierto, de minería de datos, que permite el desarrollo de modelos en un



RAPIDMINER ... también



Árbol Aplicación – Algoritmo FP Growth (Frequent Pattern Mining)





Árbol Aplicación – Algoritmo FP Growth (Frequent Pattern Mining)

- FP Growth (FP-Frequent Pattern) codifica un conjunto de datos usando una estructura de datos compacta llamada FP-tree, y extrae conjuntos de items (itemsets) frecuentes directamente desde esta estructura.
- Transacciones diferentes pueden tener varios items en común, por lo que sus caminos se pueden superponer. Mientras más caminos se superpongan, mayor compresión se puede lograr usando la estructura FP-tree. Si el tamaño del FP-tree es suficientemente pequeño para caber en memoria principal, esto permitirá extraer itemsets frecuentes desde la estructura en memoria en vez de hacer repetidas pasadas sobre los datos almacenados en disco.
- FP-Growth escanea la base de datos dos veces y usa una estructura de árbol (FP-tree) para almacenar toda la información. La raíz representa nulo, cada nodo representa un elemento, mientras que la asociación de los nodos son los conjuntos de elementos con el orden mantenido mientras se forma el árbol. El árbol FP es conciso y se utiliza para generar directamente grandes conjuntos de elementos. Una vez que se ha construido un árbol FP, utiliza un enfoque recursivo de dividir y conquistar para extraer los conjuntos de elementos frecuentes.

T.A.D. Árbol Binario de Búsqueda – Especificación (1)

Un **Árbol Binario de Búsqueda – ABB-** es un Árbol Binario en el que:

- Cada nodo contiene un campo clave (lo identifica en forma única dentro del Árbol).
- Los nodos del árbol están ordenados de tal forma que para cualquier nodo x del árbol,
 - si z es un nodo cualquiera del subárbol izquierdo de x, entonces la clave[z] < clave[x] y
 - si z es un nodo cualquiera del subárbol derecho de x, entonces clave[x] < clave[z].

Esta condición es conocida como Propiedad del Árbol Binario de Búsqueda.

T.A.D. ABB - Especificación(2)

Operaciones Abstractas

Sean A: Árbol; X,Z: claves

N	OMBRE ENCABEZADO		FUNCION	ENTRADA	SALIDA	
Cre	ear	Crear (A)	Inicializa el árbol A	A	A=()	
Ins	ertar	Insertar(A, X)	Ingresa el elemento X en el árbol	A, X	A con X como hoja, si X ∉	
$\ \ $			A, manteniéndolo como		A; Error en caso	
Ш			árbol binario de búsqueda.		contrario	
Su	primir	Suprimir(A, X)	Elimina el elemento X del árbol	A, X	A sin el nodo que contenía a	
$\ \ \ $			A, manteniéndolo como		$X, \text{ si } X \in A; \text{ Error en}$	
$\ \cdot\ $			árbol binario de búsqueda.		caso contrario	
		- (,)				
Bu	scar /	Buscar(A,X)	Localiza el nodo con clave X en	A,X	Reporta los datos asociados	
$\ \cdot\ $			el árbol A		$con X, si X \in A; Error$	
Щ	N /				en caso contrario	
Ni	vel	Nivel(A, X)	Calcula el nivel del nodo con	A, X	Reporta el nivel del nodo	
	\ \ \ \		clave X		con clave $X \text{ si } X \in A$;	
	111				Error en caso contrario	
Ho	ja 🚺	Hoja(A, X)	Evalúa si el nodo con clave X es	A, X	Verdadero si el nodo con	
			hoja		clave X es hoja, Falso	
					en caso contrario.	

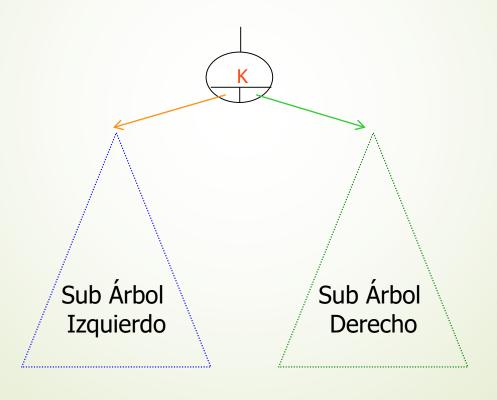
T.A.D. ABB - Especificación(3)

Operaciones Abstractas

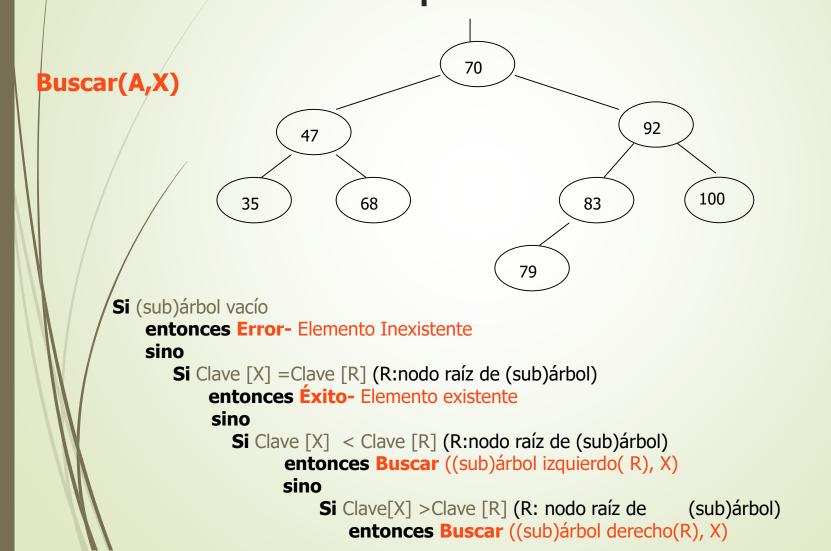
Sean A: Árbol; X,Z: claves

		/ •			, ,
	NOMBRE	ENCABEZADO	FUNCION	ENTRADA	SALIDA
H	ijo	Hijo(A, X, Z)	Evalúa si X es hijo	A, X, Z	Verdadero si el nodo X es hijo del
			(descendiente		nodo Z, Falso en caso contrario.
L		/	directo) de Z		
Pa	ndre	Padre(A, X, Z)	Recíproca de la	A, X, Z	
Ш			operación Hijo		
d	amino /	Camino(A, X, Z)	Recupera el camino del	A, X, Z	Reporta el camino de X a Z, si X es
$\ \ \ $			nodo con clave X		ancestro de Z; Error en caso
Ц			al nodo con clave Z		contrario.
A	ltura /	Altura(A)	Evalúa la altura de A	A	Reporta la altura de A.
In	Orden/	InOrden(A)	Procesa A en Inorden	A	Está sujeta al proceso que se realice
$\ \cdot\ $	W /				sobre los elementos de A
F		D (1 (A)	D A D 1	A	
P1	eOrden	PreOrden(A)	Procesa A en Preorden	A	Está sujeta al proceso que se realice
	///				sobre los elementos de A
P _i	ostOrden	PostOrden(A)	Procesa A en Postorden	A	Está sujeta al proceso que se realice
	ostorucii			11	sobre los elementos de A
	///				sourc los cicinentos de A

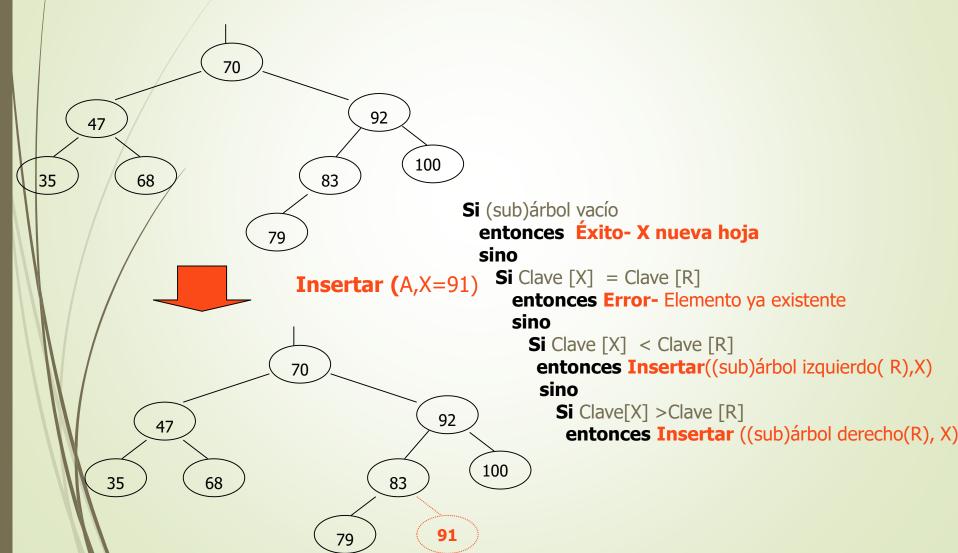
T.A.D. ABB - Representación



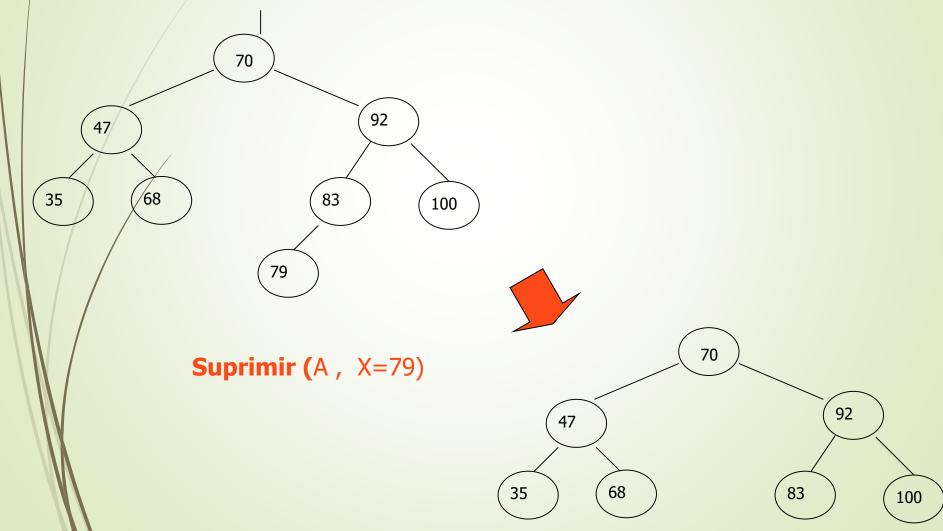
T.A.D. ABB — Construcción de Operaciones Abstractas (1)



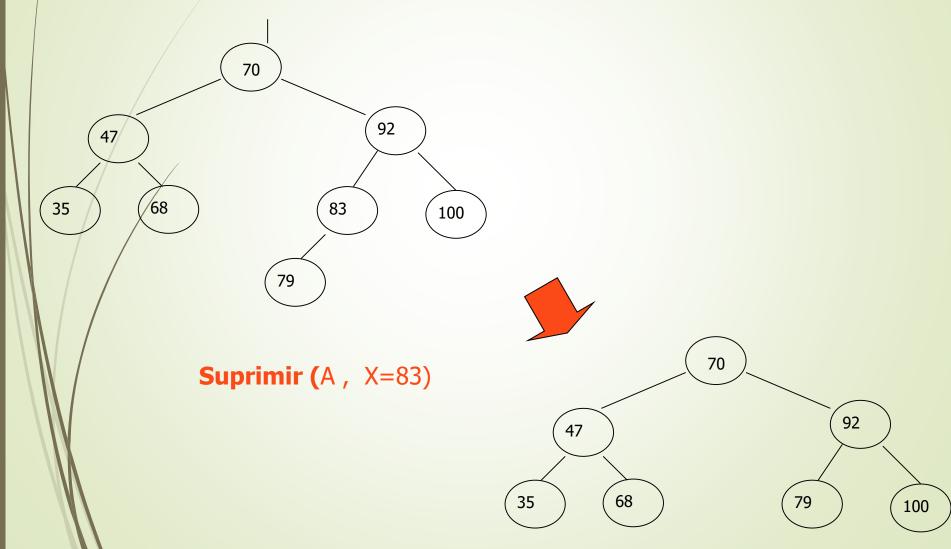
Construcción de Operaciones Abstractas (2)



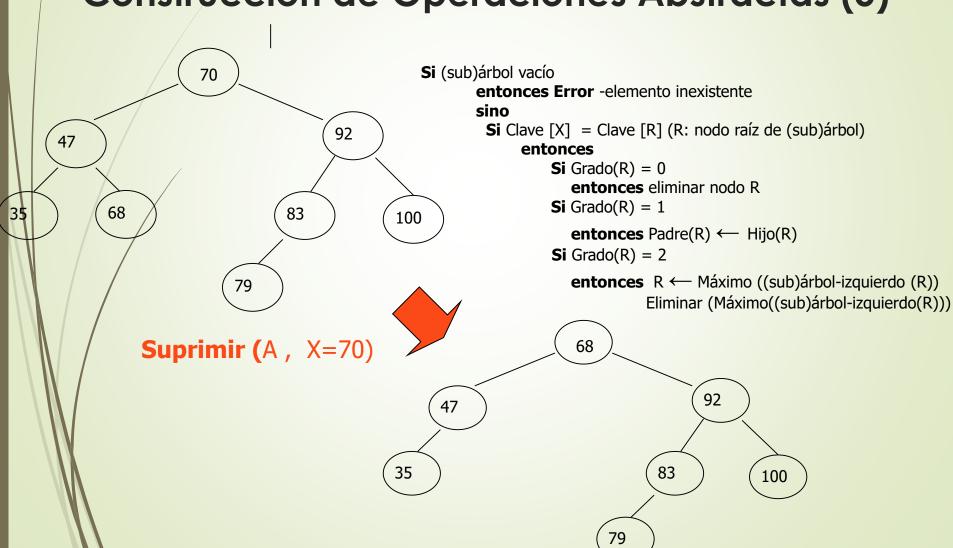
Construcción de Operaciones Abstractas (3)



Construcción de Operaciones Abstractas (4)



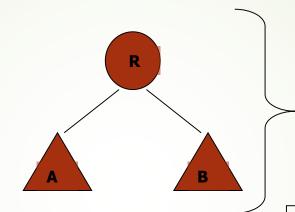
Construcción de Operaciones Abstractas (5)



Árbol Binario

Construcción de Operaciones Abstractas-Recorridos (6)

Recorridos



R: raiz

A: Sub Árbol izq de R

B: Sub Árbol der de R

Preorden (Árbol)

Si (Arbol no vacío)

entonces

Procesar nodo R

Preørden ((Sub)Árbol izq de R)

Preorden ((Sub)Árbol der de R)

Inorden (Árbol)

Si (Árbol no vacío)

entonces

Inorden((Sub) Árbol izq de R)

Procesar nodo R

Inorden((Sub) Árbol der de R)

Postorden (Arbol)

Si (Árbol no vacío)

entonces

Árbol

Postorden((Sub) Árbol izq

de R)

Postorden((Sub) Árbol der

de R)

Procesar nodo R

T.A.D. ABB - Aplicación

Indice de referencias cruzadas

un

SALIDA ENTRADA al de editor El editor de lenguaje C 2, 3 entorno presenta un entorno el similar al entorno 1, 4 Lenguaje de lenguaje Pascal` Pascal 2 3 2 presenta similar

Árbol Binario – Tipos

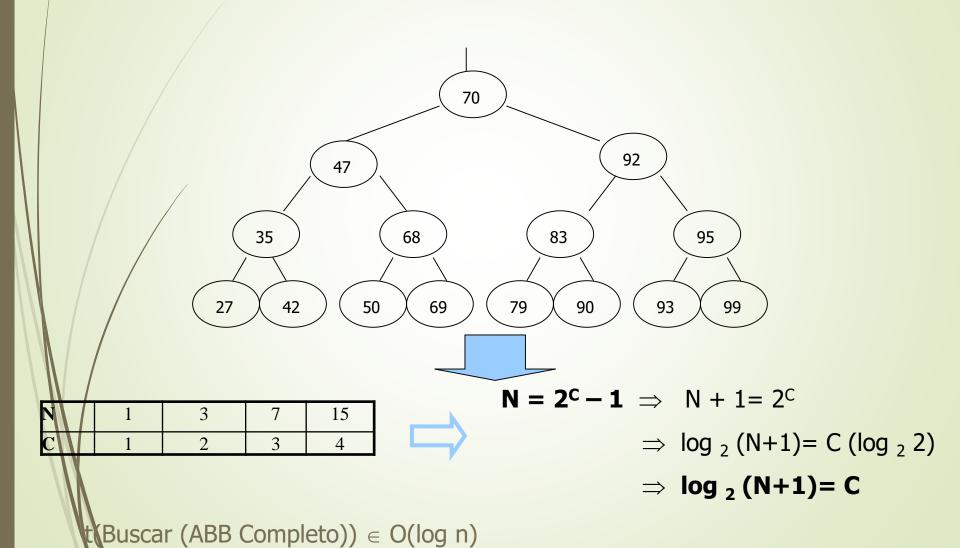
¿Cuántas comparaciones se realizan en una **búsqueda**, en el peor de los casos, en un **ABB** ?

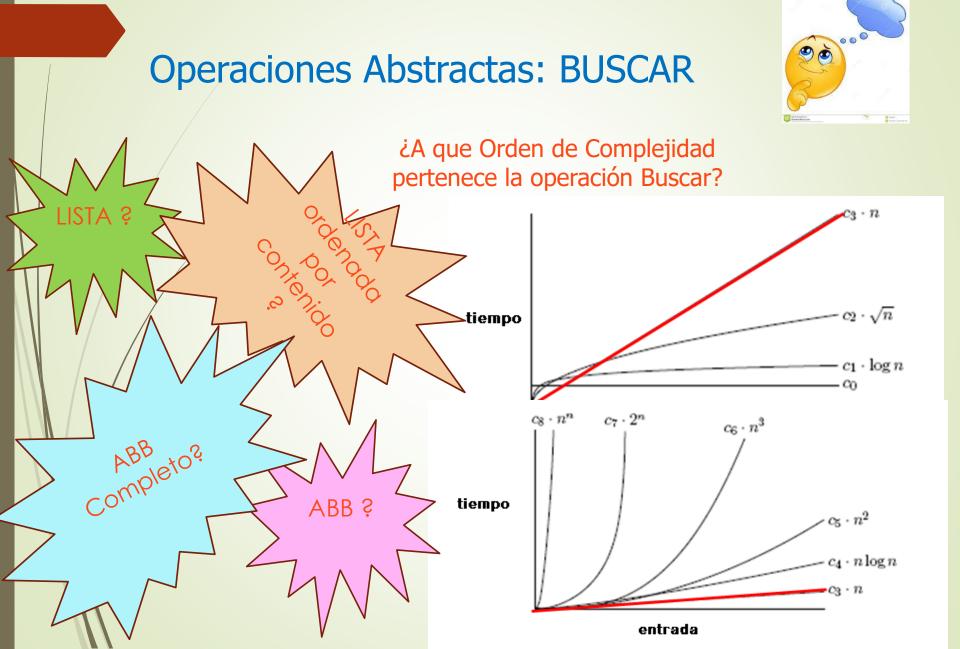
Las operaciones Insertar, Suprimir y Buscar se realizan en *h* accesos como máximo, siendo *h* la altura del árbol binario.

- Árbol Binario Relleno: Es aquel árbol en el que todo nodo o bien es una hoja, o tiene dos hijos.
- Árbol Binario Completo: Es un árbol binario relleno en el que todas las hojas están en el mismo nivel.

¿Cuántas comparaciones se realizan en una **búsqueda**, en el peor de los casos, en un **ABB Completo**?

ABB Completo





Arbol Balanceado- AVL

Un árbol es *perfectamente equilibrado*, si para cada nodo los números de nodos en sus subárboles izquierdo y derecho difieren cuanto más en uno.

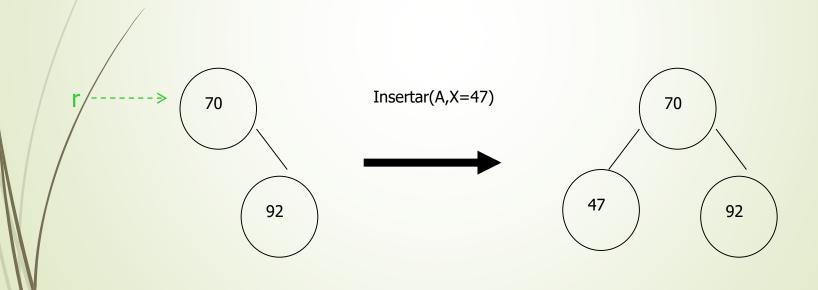
Un árbol está *balanceado* si y solo si en cada nodo las alturas de sus dos subárboles difieren a lo máximo en uno.

Los árboles balanceados reciben el nombre de **Árboles AVL** por ser Adelson-Velski y Landis quienes propusieron esta definición de equilibrio.

T.A.D. Arbol Balanceado – Construcción de operaciones abstractas (1)

Inserción en un Árbol Balanceado

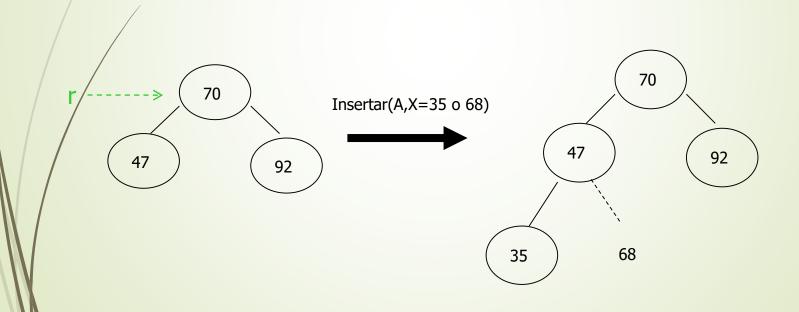
a) Si altura(I(r)) < altura(D(r)) y X se inserta en I(r)



T.A.D. Arbol Balanceado – Construcción de operaciones abstractas (2)

Inserción en un Árbol Balanceado

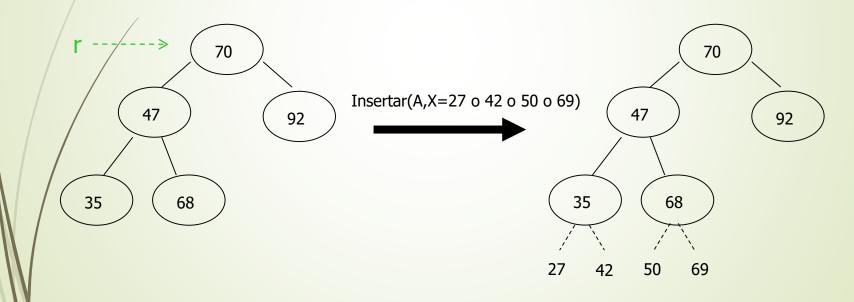
b) Si altura(I(r)) = altura(D(r)) y X se inserta en I(r),



T.A.D. Arbol Balanceado – Construcción de operaciones abstractas (3)

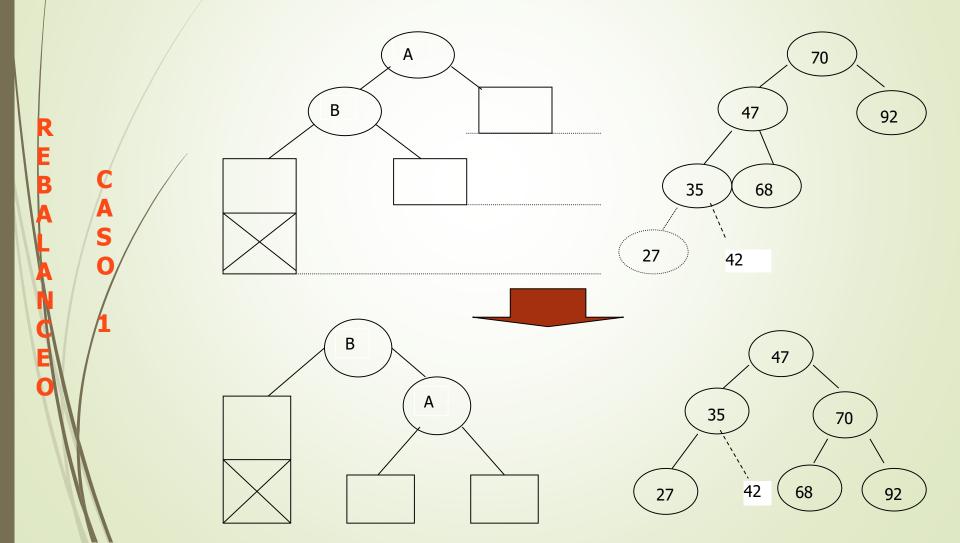
Inserción en un Árbol Balanceado

c) Si altura(I(r)) > altura(D(r)) y X se inserta en I(r)

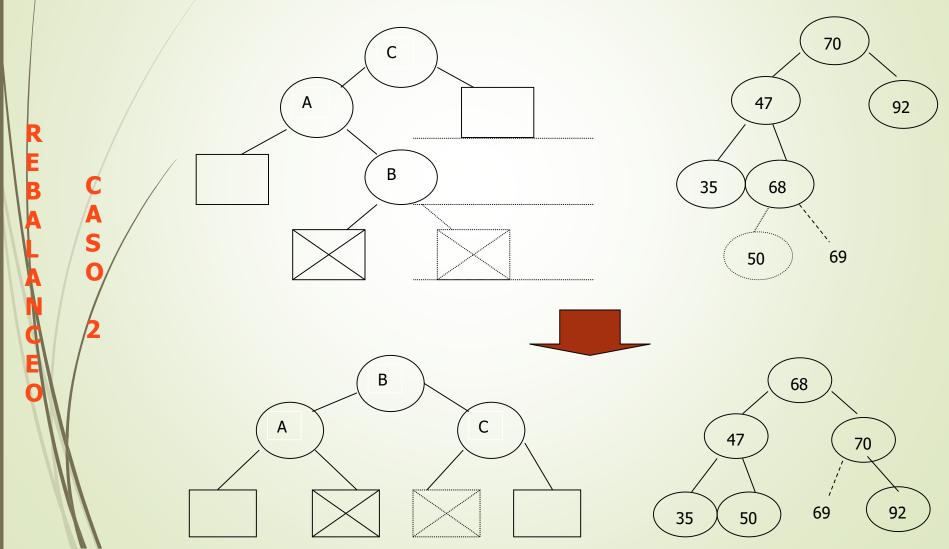


REBALANCEAR!!

T.A.D. Arbol Balanceado – Construcción de operaciones abstractas (4)



T.A.D. Arbol Balanceado – Construcción de operaciones abstractas (5)



Arbol Multicamino

Construcción y mantención de árboles de búsqueda a gran escala, que se almacenan en memoria secundaria.

Almacenar datos de 1 millón de elementos árbol balanceado $Log_10 = 20^6 comp.$



acceso a disco por cada comp

20 accesos a disco

Se incorpora un tipo particular de árbol multicamino cantidad de accesos en el peor de los casos sería: $Log_{100} 10 = 3$ accesos

Cada página (salvo una) contiene entre *n* y 2*n* nodos para determinada constante n.

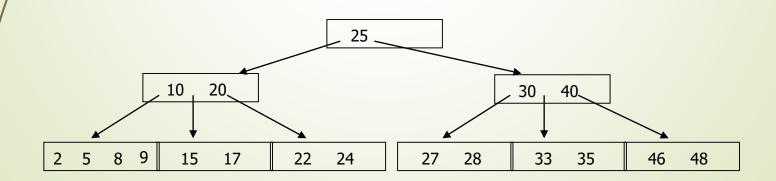
De ahí que, en un árbol con N elementos y un tamaño máximo de página de 2n nodos por página, en el peor caso requiere log N accesos de página.

Árbol B, árbol multicamino de orden n:

- 1) Cada página contiene a lo sumo 2n elementos (claves).
- 2) Cada página, excepto la pagina raíz, contiene n elementos por lo menos.
- 3) Cada página es una página de hoja, o sea que no tiene descendientes, o tiene m+1 descendientes, donde m es su número de claves en esa página $(n < \neq m < = 2n)$.
- 4/) Todas las páginas hoja aparecen al mismo nivel.

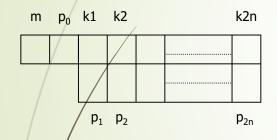
T.A.D. Árbol B de orden 2 Especificación

- 1) Cada página contiene a lo sumo 4 (2*2) elementos (claves).
- 2) Cada página, excepto la pagina raíz, contiene 2 elementos por lo menos.
- 3) Cada página es una página de hoja, o sea que no tiene descendientes, o tiene m+1 descendientes, donde m es su número de claves en esa página (2<=m<=4).
- 4) Todas las páginas hoja aparecen al mismo nivel.



T.A.D. Árbol B Representación

Estructura de la Página



m : cantidad de claves en la página

ki: clave; 1 <= i <= m

p0 : dirección de la página que contiene claves

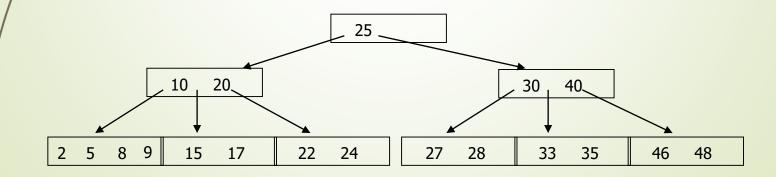
menores que k1

pi : dirección de la página que contiene claves

mayores que ki y menores que ki+1

pm : dirección de la página que contiene claves

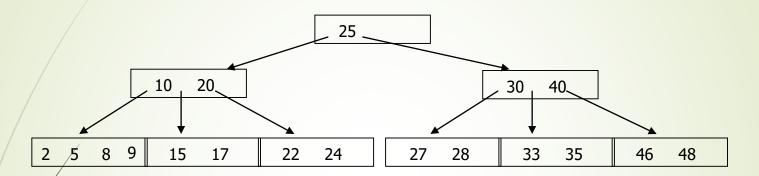
mayores que km



Estructuras de Datos y Algoritmos

T.A.D. Árbol B

Operaciones Abstractas: Buscar



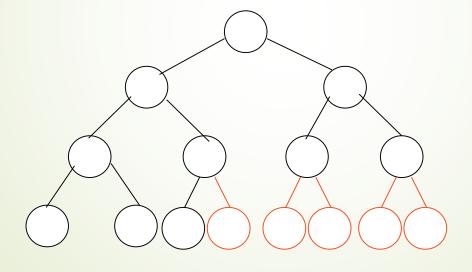
¿Cómo se realiza la búsqueda de una clave **X**, *en* una página dada? ¿Si la clave X no se encuentra en una página, cómo continúa la búsqueda?

- 1) km < X;
- 2) X< k1;
- 3) ki < X < ki+1 (1 <= i < m)

entonces la búsqueda continúa por la página apuntada por p_m entonces la búsqueda continúa por la página apuntada por p₀ entonces la búsqueda continúa por la página apuntada por pi

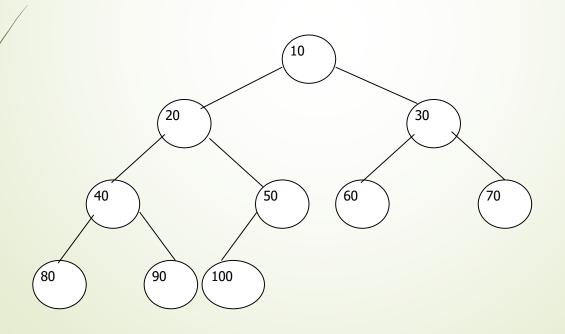
Árbol Binario Semicompleto

Árbol Binario Semicompleto: Un Árbol Binario Semicompleto de **n** nodos, se forma a partir de un árbol binario completo de **n+q** nodos, quitando las **q** hojas extremo derechas del árbol binario completo.



T.A.D.Montículo Binario - Especificación

Montículo Binario: Un *Montículo Binario* es un árbol binario semicompleto en el que el valor de clave almacenado en cualquier nodo es menor o igual que el valor de clave de sus hijos.



Montículos Binarios Colas de Prioridad



Los valores —claves- que representan prioridades deben interpretarse de la siguiente manera: a menor valor-mayor prioridad, por lo que la máxima prioridad se encuentra en la raíz del árbol.

T.A.D.Montículo Binario - Especificación

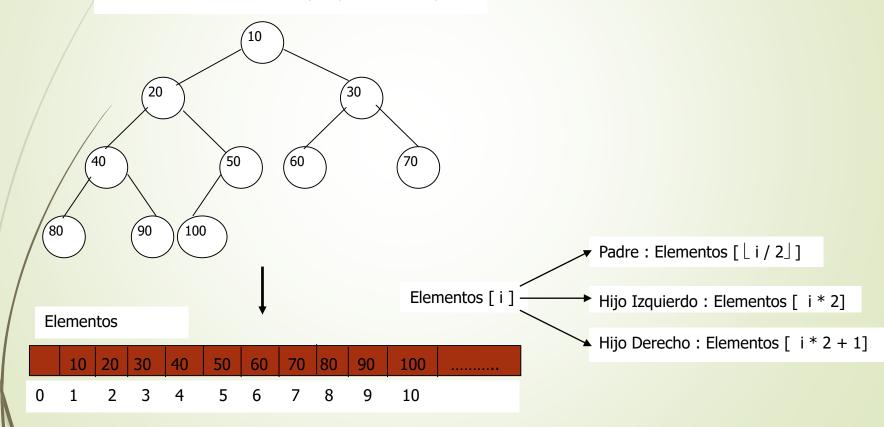
Operaciones Abstractas

M: Montículo Binario y X: Clave

NOMBRE	ENCABEZADO	FUNCIÓN	ENTRADA	SALIDA
Insertar	Insertar(M, X)	Ingresa el elemento X al montículo M	M, X	M con el nuevo elemento
Eliminar_Mínimo	Eliminar_Mínimo (M, X)	Suprime del montículo M el elemento de máxima prioridad- mínimo valor de clave	М	M y X: elemento de máxima prioridad

T.A.D. Montículo Binario - Representación

Montículo Binario desde una perspectiva conceptual



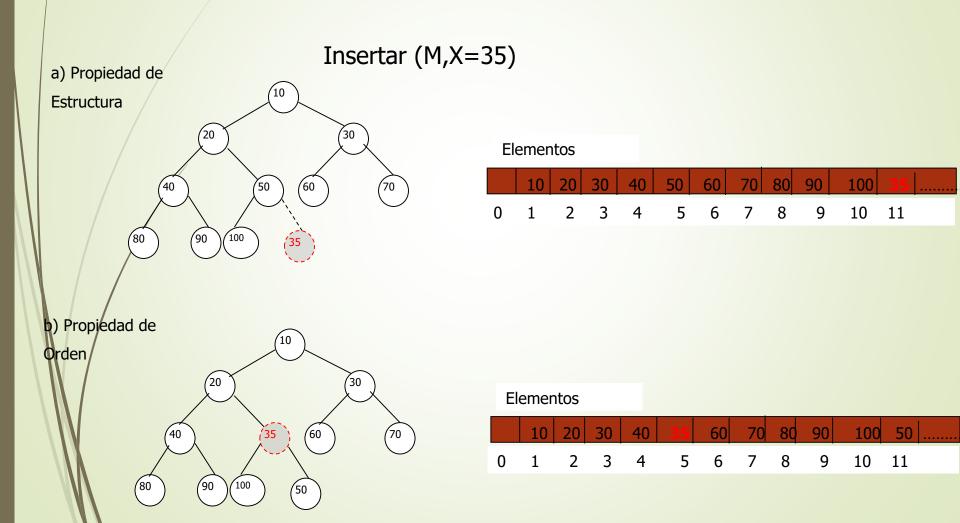
Montículo Binario en su almacenamiento

T.A.D.Montículo Binario – Construcción de operaciones abstractas (1)

Tanto la operación *Insertar* como *Eliminar_Mínimo*, deben garantizar que en el Montículo Binario se mantengan las siguientes dos propiedades:

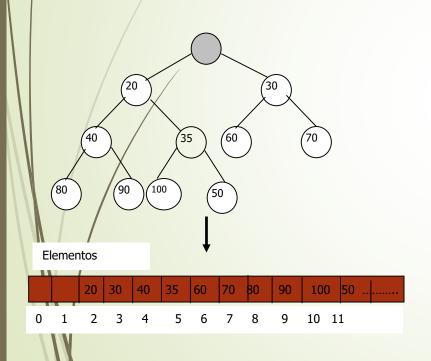
- ➢ Propiedad de Estructura: el objeto de datos debe ser un árbol binario semicompleto.
- ➤ **Propiedad** de **Orden** : el valor de clave almacenado en cualquier nodo debe ser menor o igual que el valor de clave de sus hijos.

T.A.D.Montículo Binario – Construcción de operaciones abstractas (2)



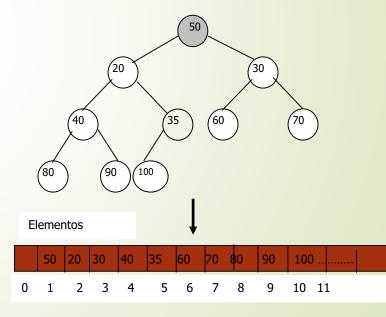
T.A.D.Montículo Binario – Construcción de operaciones abstractas (3)

Eliminar_Mínimo (M, X)



a) Propiedad de Estructura

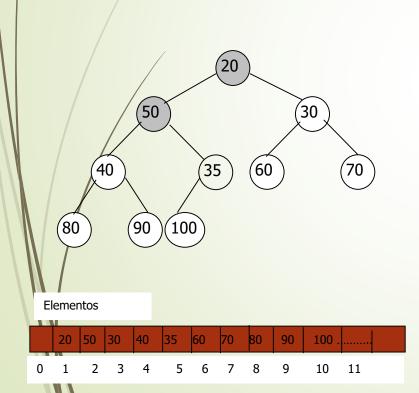


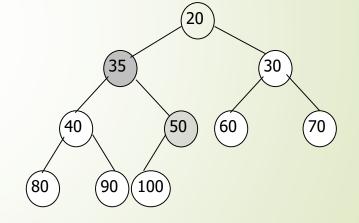


T.A.D.Montículo Binario -Construcción de operaciones abstractas (4)

Eliminar_Mínimo (M, X)

b) Propiedad de Orden





Elementos

	20	35	30	40	50	60	70	80	90	100 .		
0	1	2	3	4	5	6	7	8	9	10	11	