Strutture dati - Parte 2

Dipartimento di Elettronica, Informazione e Bioingegneria Politecnico di Milano

22 maggio 2024

Alberi

Una struttura dati flessibile

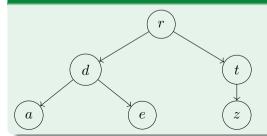
- Una struttura dati versatile è il cosiddetto albero
- Ne abbiamo già fatto uso informalmente (limite inferiore complessità dell'ordinamento per confronto)
- In estrema sintesi:
 - Un albero è costituito da un insieme di nodi e uno di archi che li collegano
 - Ogni nodo ha al più un arco entrante, ma un numero arbitrario di archi uscenti
- Gli alberi sono una rappresentazione efficiente per insiemi di dati ordinati

Alberi

Definizione

Un albero A è una coppia (V, E) dove V è un insieme di nodi e E un insieme di archi (coppie di nodi ordinate). Ogni nodo può apparire un'unica volta come destinazione di un arco (= secondo elemento della coppia). Non sono possibili cicli.

Graficamente



- $\mathbf{V} = \{r, a, d, e, t, z\}$
- $\mathbf{E} = \{(r, d), (r, t), (d, a), (d, e), (t, z)\}$

Alberi

Nomenclatura

- Radice: É l'unico nodo dell'albero privo di un arco entrante
- Foglia: Un nodo senza archi uscenti
- Padre (o genitore): di un nodo n: il nodo da cui l'arco entrante in n ha origine
- ullet Figlio (o discendente): di un nodo n: il nodo in cui uno degli archi uscenti da n termina
- Un albero in cui ogni nodo ha al più due figli è detto albero binario
- Livello: distanza, in numero di archi, di un nodo dalla radice
- Albero completo: un albero in cui tutti i livelli hanno tutti i nodi tranne l'ultimo
 - i nodi dell'ultimo livello sono impacchettati da sinistra

Alberi binari

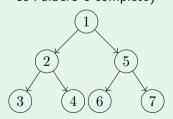
In pseudocodice

- Ci occuperemo di alberi binari (= ogni padre ha 2 figli)
- É utile dare una definizione ricorsiva di albero:
 - Un albero è formato da un nodo radice a cui sono collegati due alberi, il sottoalbero destro e quello sinistro
 - Un albero può essere vuoto (NIL)
- Le azioni sull'albero indicizzeranno i nodi con una chiave
 - Simile a quanto fatto per le tabelle hash, la chiave è un intero
- Dato un nodo A:
 - A.left è il riferimento al figlio sinistro, A.right al destro
 - A.p è il riferimento al padre, A.key è la chiave
- ullet Ogni albero A ha un riferimento A.root alla radice
 - A.p è NIL solo per la radice

Stoccaggio con struttura dati implicita

Alberi binari stoccati con un vettore

- Un albero può essere materializzato in memoria naturalmente con una struttura basata su puntatori
- Alternativamente è possibile utilizzare un vettore per contenere le chiavi (efficiente se l'albero è completo)



- La radice dell'albero è stoccata nella prima posizione del vettore
- Dato un nodo contenuto in posizione i il suo figlio sx è in posizione 2i+1, il dx in 2i+2 (contando da 0 le pos.
- Il padre del nodo stoccato in posizione i (se esiste) si trova in posizione $\lfloor \frac{i-1}{2} \rfloor$

Alberi binari

Visita di un albero

- Su di un albero è possibile effettuare operazioni di inserimento, ricerca e cancellazione di nodi come sulle altre strutture dati
- L'operazione caratteristica degli alberi è il cosiddetto *attraversamento* o *visita* per enumerare le chiavi contenute
- La definizione naturale degli algoritmi di visita è ricorsiva
- Il fattore discriminante tra le diverse strategie è l'*ordine* in cui i nodi vengono visitati

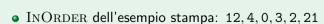
Visita di un albero

Visita in-ordine (in-order)

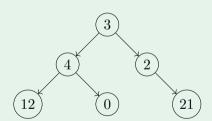
 Nella visita in ordine si visita prima il sottoalbero sx, quindi la radice, infine il sottoalbero dx

InOrder(T)

- 1 INORDER(T.left)
- 2 Print(T.key)
- 3 INORDER(T.right)
- 4 return



• Complessità: $\Theta(n)$, tocca una sola volta ogni nodo



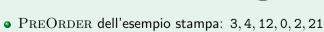
Visita di un albero

Visita anticipata (preorder)

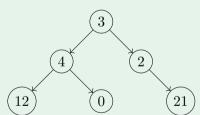
 Nella visita in ordine si visita prima la radice, quindi il sottoalbero sx, infine il sottoalbero dx

PreOrder(T)

- 1 Print(T.key)
- 2 PREORDER(T.left)
- 3 PREORDER(T.right)
- 4 return



• Complessità: $\Theta(n)$, tocca una sola volta ogni nodo



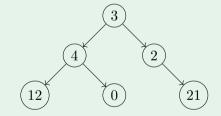
Visita di un albero

Visita posticipata (postorder)

• Nella visita in ordine si visita prima il sottoalbero sx, poi il sottoalbero dx e infine la radice

PostOrder(T)

- 1 PostOrder(T.left)
- 2 PostOrder(T.right)
- 3 Print(T.key)
- 4 return



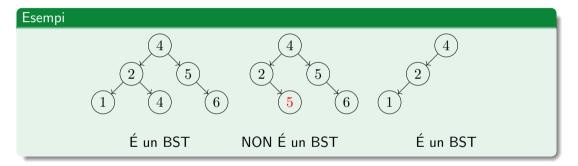
- POSTORDER dell'esempio stampa: 12, 0, 4, 21, 2, 3
- Complessità: $\Theta(n)$, tocca una sola volta ogni nodo

|Alberi binari di ricerca (BST)

Definizione

- Uno degli usi più comuni degli alberi binari è utilizzare quelli per cui è valida una data relazione tra le chiavi
- ullet Un albero binario è un albero binario di ricerca se per ogni suo nodo x valgono:
 - Se y è contenuto nel sottoalbero sinistro di x, $y.key \le x.key$
 - Se y è contenuto nel sottoalbero destro di x, $y.key \ge x.key$
- Inserimenti e cancellazioni devono preservare la proprietà
- Una visita in-ordine del BST stampa le chiavi in ordine

Alberi binari di ricerca (BST)



Nota

• Cambiare la condizione sui valori di chiave rimuovendo la possibilità che siano uguali rende gli elementi del BST unici

Ricerca

 La struttura dei BST li rende naturali candidati per una ricerca efficace degli elementi per chiave

```
\begin{array}{ll} \operatorname{RICERCA}(T,x) \\ 1 & \text{if } T = \operatorname{NIL} \text{ or } T.key = x.key \\ 2 & \text{return } T \\ 3 & \text{if } T.key < x.key \\ 4 & \text{return } \operatorname{RICERCA}(T.right,x) \\ 5 & \text{else return } \operatorname{RICERCA}(T.left,x) \\ 6 & \end{array}
```

- Complessità: $\mathcal{O}(h)$ con h l'altezza dell'albero
- Nel caso ottimo (albero "ben bilanciato") diventa $\mathcal{O}(\log(n))$
- Nel caso pessimo (albero degenere in lista) è $\mathcal{O}(n)$

Minimo e massimo

• L'elemento con chiave minima (massima) è quello più a sinistra (destra) del BST

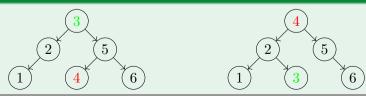
```
\begin{array}{lll} \operatorname{Min}(T) & \operatorname{Max}(T) \\ 1 & \operatorname{cur} \leftarrow T & 1 & \operatorname{cur} \leftarrow T \\ 2 & \text{while } \operatorname{cur.left} \neq \operatorname{NIL} & 2 & \text{while } \operatorname{cur.right} \neq \operatorname{NIL} \\ 3 & \operatorname{cur} \leftarrow \operatorname{cur.left} & 3 & \operatorname{cur} \leftarrow \operatorname{cur.right} \\ 4 & \operatorname{return } \operatorname{cur} & 4 & \operatorname{return } \operatorname{cur} \end{array}
```

ullet Complessità: $\mathcal{O}(h)$ con h l'altezza dell'albero

Successore

- Il successore di un elemento x è l'elemento y con la più piccola chiave y.key > x.key presente nel BST
- Nel cercarlo sono possibili due casi:
 - 1 Il sottoalbero dx di x non è vuoto: il successore è il minimo di quel sottoalbero
 - Il sottoalbero dx di x è vuoto: il successore è il progenitore più prossimo a x per cui x appare nel suo sottoalbero sx

Esempi: Successore di 3



Successore

• Lo pseudocodice per la ricerca del successore è il seguente:

```
\begin{array}{lll} {\rm SUCCESSORE}(x) \\ 1 & \mbox{if } x.right \neq {\rm NIL} \\ 2 & \mbox{return } {\rm Min}(x.right) \\ 3 & y \leftarrow x.p \\ 4 & \mbox{while } y \neq NIL {\rm \ and \ } y.right = x \\ 5 & x \leftarrow y \\ 6 & y \leftarrow y.p \\ 7 & \mbox{return } y \end{array}
```

- Complessità nel caso 1: la stessa del calcolo del minimo: $\mathcal{O}(h)$
- Complessità nel caso 2: caso pessimo, x è la foglia più distante dalla radice, $\mathcal{O}(h)$

Inserimento – Struttura

- L'inserimento di un nuovo elemento deve rispettare la proprietà fondamentale del BST
- Assunzione: non vogliamo che il BST contenga duplicati
- Idea: Cerco l'elemento che voglio inserire nel BST, non lo trovo, lo inserisco al posto del NIL trovato.
- Unica accortezza rispetto al codice della ricerca: tenere traccia dell'ultimo nodo non-NIL per poter inserire correttamente l'elemento

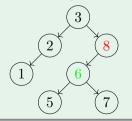
Inserimento – Pseudocodice

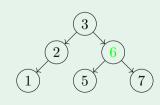
```
Inserisci(T, x)
     pre \leftarrow \texttt{NIL}
     cur \leftarrow T.root
      while cur \neq NIL
            pre \leftarrow cur
            if x.key < cur.key
 6
                   cur \leftarrow cur.left
            else cur \leftarrow cur.right
      x.p \leftarrow pre
      if pre = NIL
10
            T.root \leftarrow x
      elseif x.key < pre.key
12
           pre.left \leftarrow x
      else pre.right \leftarrow x
```

- Le righe 3–7 effettuano la ricerca della posizione di inserimento nell'albero
- Le righe 8–13 effettuano l'inserimento vero e proprio
- Complessità: la stessa della ricerca $\mathcal{O}(h)$ più una porzione a tempo costante (inserimento)

Cancellazione – Struttura – 1

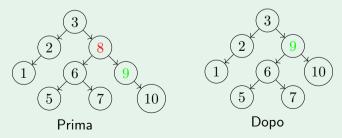
- La strategia di cancellazione di un elemento da un BST dipende dal numero di figli dell'elemento in questione
- Caso 1 L'elemento non ha figli: è sufficiente eliminarlo dall'albero deallocandolo e impostando il puntatore del padre a NIL
- Caso 2 L'elemento ha un figlio: L'elemento viene sostituito dal figlio nel suo ruolo nell'albero





Cancellazione – Struttura – 2

- Caso 3 L'elemento ha due figli: copio il valore del suo successore su di esso ed elimino il successore
- Il successore s di un elemento con due figli x non ha mai il figlio sx f: si avrebbe x.key < f.key < s.key, ma questo è impossibile per definizione di successore



Cancellazione - Pseudocodice

```
Cancella(T, x)
      if x.left = \mathtt{NIL} or x.right = \mathtt{NIL}
            da\ canc \leftarrow x
      else da\_canc \leftarrow Successore(x)
      if da\_canc.left \neq \texttt{NIL}
            sottoa \leftarrow da\_canc.left
      else sottoa \leftarrow da\_canc.right
      if sottoa \neq NIL
            sottoa.p \leftarrow da\_canc.p
      if da_{-}canc.p = NIL
10
            T.root \leftarrow sottoa
      elseif da\_canc = da\_canc.p.left
12
            da\_canc.p.left \leftarrow sottoa
      else da\_canc.p.right \leftarrow sottoa
      if da\_canc \neq x
15
            x.key \leftarrow da\_canc.key
      Free(da\_canc)
```

- Le righe 1–3 individuano il nodo da cancellare
- Le righe 4–8 individuano il sottoalbero da spostare e correggono il riferimento al padre
- Le righe 9–13 correggono il riferimento del padre
- Le righe 14–15 copiano il valore della chiave

Analisi di complessità

Sommario

- Tutte le operazioni sono $\mathcal{O}(h)$ con h l'altezza del BST
- Nel migliore dei casi $h = \log(n)$ (albero completo o quasi completo), nel peggiore h = n (lista)
- É critico per avere buone prestazioni mantenere il BST il più possibile vicino al caso ottimo
- Si può dimostrare che l'altezza attesa di un BST è $\mathcal{O}(\log(n))$ se le chiavi inserite hanno distribuzione uniforme
- Volendo un metodo deterministico ci serve una definizione di albero ben bilanciato

Albero bilanciato

Vicinanza all'albero completo

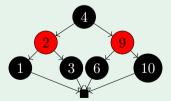
- Intuitivamente, vogliamo che la distanza delle foglie dalla radice sia limitata superiormente, per tutte le foglie
- Una definizione operativa (Adelson-Velskii e Landis, 1962)
 - \bullet Un albero è bilanciato se, per ogni nodo, le altezze dei due sottoalberi differiscono al più di 1
- Adelson-Velskii e Landis proposero, insieme alla definizione, una modifica ai BST ed ai metodi per accedervi in grado di tenerli bilanciati (alberi AVL)
- Vediamo un'ottimizzazione degli alberi AVL che sacrifica parte del bilanciamento per ottenere inserimenti/cancellazioni più efficienti: gli alberi rosso-neri (red-black trees, RB-trees)
 - Sia la perdita in bilanciamento, che il guadagno sono costanti

Struttura e definizione

- Un albero rosso-nero è un BST i cui nodi sono dotati di un attributo aggiuntivo, detto $colore \in \{rosso, nero\}$, e soddisfacente le seguenti 5 proprietà:
 - Ogni nodo è rosso o nero
 - La radice è nera
 - Le foglie sono nere
 - I figli di un nodo rosso sono entrambi neri
 - Per ogni nodo dell'albero, tutti i cammini dai suoi discendenti alle foglie contenute nei suoi sottoalberi hanno lo stesso numero di nodi neri
- ullet Chiamiamo, per comodità, altezza nera (black height) di un nodo x il valore bh(x) pari al numero di nodi neri, escluso x se è il caso, nel percorso che va da x alle foglie

Convenzioni

- I dati sono mantenuti unicamente nei nodi interni, le foglie sono tutte NIL
- Per semplicità, tutte le foglie sono fisicamente rappresentate da un singolo nodo, il cui unico riferimento T.nil è conservato nella struttura dati
- Il padre del nodo radice punta anch'esso a T.nil



Azioni sugli alberi rosso-neri

- Tutte le operazioni che non vanno a modificare la struttura dell'albero sono identiche ai BST: RICERCA, MIN, MAX, SUCCESSORE, PREDECESSORE
- Le operazioni di INSERISCI e CANCELLA hanno necessità di mantenere le proprietà degli alberi rosso-neri
 - Idea di massima: opero come se si trattasse di un BST generico, dopodichè compenso le eventuali violazioni
- É necessario essere in grado di ri-bilanciare l'albero con modifiche solamente locali (no ricostruzione dell'albero)

Teorema (Proprietà di buon bilanciamento)

Un albero RB con n nodi interni ha altezza massima $2\log(n+1)$

Dimostrazione - 1

- Dim. che un sottoalbero con radice x ha almeno $2^{bh(x)}-1$ nodi interni, per induzione sull'altezza del sottoalbero
 - Caso base (altezza 0): x è una foglia, il sottoalbero contiene almeno $2^{bh(x)} 1 = 2^0 1 = 0$ nodi interni
 - Passo: dato x, entrambi i suoi figli hanno altezza nera bh(x) o bh(x)-1. Dato che l'altezza dei figli è minore di quella di x (per hp. ind.) i loro sottoalberi hanno almeno $2^{bh(x)-1}-1$ nodi interni. L'albero radicato in x contiene quindi almeno $2^{bh(x)-1}-1+2^{bh(x)-1}-1+2=2^{bh(x)}$ nodi

Dimostrazione.

- Per la proprietà 4, almeno metà dei nodi su un qualunque percorso radice-foglia sono neri
- L'altezza nera della radice è almeno $\frac{h}{2}$; per quanto detto prima il sottoalbero radicato in essa contiene almeno $2^{\frac{h}{2}}-1$ nodi
- $n \ge 2^{\frac{h}{2}} 1$, risolvendo per $h \to h \le 2\log(n+1)$

Conseguenze

- Le operazioni che restano invariate tra RB-trees e BST (RICERCA, MAX, MIN, SUCCESSORE) sono $\mathcal{O}(2\log(n+1))$
- Se riesco a riparare le violazioni in maniera efficiente ho anche INSERISCI e CANCELLA) in $\mathcal{O}(2\log(n+1))$

Rotazione

Operazione locale a due nodi di un BST che cambia il livello a cui sono situati due nodi senza violare la proprietà BST

LeftRotate
$$(T,x) \rightarrow$$

$$l(\rho) = i - 1$$

$$l(x) = i$$

$$l(\alpha) = i + 1$$

$$l(y) = i + 1$$

$$l(\beta) = i + 2$$

$$l(\gamma) = i + 2$$

$$\leftarrow RIGHTROTATE(T, Y)$$

$$l(\rho) = i - 1$$

 $l(x) = i + 1$
 $l(\alpha) = i + 2$
 $l(y) = i$
 $l(\beta) = i + 2$
 $l(\gamma) = i + 1$

Rotazioni

Pseudocodice

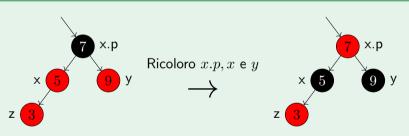
```
LeftRotate(T, x)
    u \leftarrow x.right // Calcola u
    x.right \leftarrow y.left // Sposta il sottoalbero \beta
     if y.left \neq T.nil // Sistema il riferimento a p della radice di \beta
           u.left.p \leftarrow x
     y.p \leftarrow x.p // Sistema il riferimento al padre del nodo y
     if x.p = T.nil // Sistema il riferimento al figlio del padre di y
           T.root \leftarrow u
     elseif x = x.p.left
           x.p.left \leftarrow y
    else x.p.right \leftarrow y
    y.left \leftarrow x // Aggancia x a sinistra di y
12 x.p \leftarrow y
```

Inserimento

- L'inserimento procede ad inserire il nuovo elemento come se l'albero fosse un semplice BST salvo:
 - ullet Assegnare il valore dei sottoalberi del nodo a T.nil al posto di NIL se viene inserito come una foglia
 - Assegnare il valore del genitore del nodo a T.nil al posto di NIL se il nodo è inserito come radice
 - Colorare il nodo appena inserito di rosso
- Possono essere violate la proprietà 4 (i figli di un nodo rosso sono entrambi neri) e la 2 (la radice è nera)
- RIPARARBINSERISCI(z), dato il nodo inserito z:
 - Con z.p figlio sx del nonno di z: 3 casi a seconda del colore dello "zio" e della posizione di z rispetto a z.p
 - Simmetrica per il caso in cui il padre è figlio dx del nonno

RIPARARBINSERISCI(z)

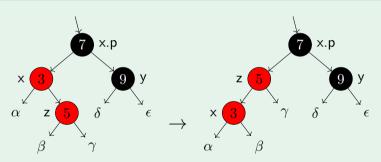
Caso 1 - lo "zio" y è rosso, posizione di z irrilevante, z.key=3



- ullet Successivamente chiamo RIPARARBINSERISCI(x.p): x.p potrebbe avere un padre rosso
- ullet Se x.p è la radice, posso colorarla di nero senza problemi

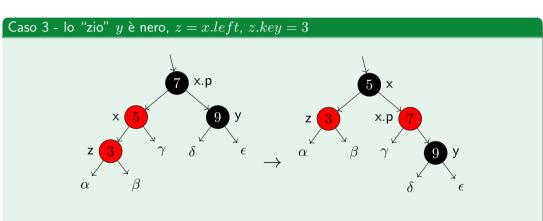
RIPARARBINSERISCI(z)

Caso 2 - lo "zio" y è nero, z=x.right, z.key=5



• Effettuiamo LeftRotate(T,x): ora la riparazione va effettuata su z, con z.left=x rosso (caso successivo)

RIPARARBINSERISCI(z)



ullet Scambiamo i colori di x e x.p ed eseguiamo $\operatorname{RIGHTROTATE}(T,p.x)$

Operazioni su alberi RB

RiparaRBInserisci – Pseudocodice

```
RIPARARBINSERISCI(T, z)
      while z.p.color = red
           if z.p = z.p.p.left
                 y \leftarrow z.p.p.right
                 if y.color = red \# \mathsf{Caso} \ 1
                      z.p.color \leftarrow black
                      y.color \leftarrow black
                      z.p.p.color \leftarrow red
                      z \leftarrow z.p.p
                 else if z=z.p.right // Caso 2, ruotare z
10
                           z \leftarrow z.p
11
                            LeftRotate(z)
12
                      z.p.color \leftarrow black // Caso 3
13
                      z.p.p.color \leftarrow red
                      RIGHTROTATE(Z.P.P)
14
15
           else // come le righe 3-14, scambiando right con left
      T.root.color \leftarrow black
```

Alberi rosso-neri - Inserimento

Analisi di complessità

- Nei casi 2 e 3 la procedura RIPARARBINSERISCI(z) deve solo effettuare un cambio di colori locale e 2 o 1 rotazioni
 - Tutte queste operazioni sono $\mathcal{O}(k)$
- Nel caso 1 continua analizzando il nonno del nodo corrente
 - Caso pessimo: il ciclo itera numero di volte pari a metà dell'altezza dell'intero albero
- L' intera riparazione prende al più $\mathcal{O}(log(n)) \to$ l'inserimento, comprensivo di riparazione, in alberi RB è $\mathcal{O}(\log(n))$

Cancellazione

- La cancellazione procede a cancellare il nuovo elemento come se l'albero fosse un semplice BST salvo:
 - ullet l' uso di T.nil al posto di NIL
 - invocare la procedura che ripara eventuali violazioni delle proprietà RB
- Nel caso sia eliminato un nodo rosso, non è necessario alcun cambiamento (non sono possibili violazioni delle proprietà)
- ullet Se il nodo eliminato è nero, RIPARARBCANCELLA(x), dato il nodo x presente al posto di quello cancellato z ripristina le proprietà

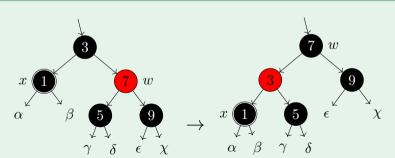
Struttura di RIPARARBCANCELLA(x)

- I 5 casi della procedura RIPARARBCANCELLA(x), con x figlio sx di x.p (procedura simmetrica se figlio dx) sono:
- ullet Caso 0: x è rosso
- Se x è nero
 - Caso 1: Il suo fratello è rosso
 - Se il suo fratello è nero
 - Caso 2: x ha entrambi i nipoti neri
 - Caso 3: x ha entrambi il nipote sinistro rosso
 - Caso 4: x ha entrambi il nipote destro rosso
- La riparazione da effettuare deve sistemare il fatto che x ha un "colore nero aggiuntivo oltre al proprio": il nero che ha "ereditato" dal nodo cancellato

Cancellazione - Caso 0: x è rosso

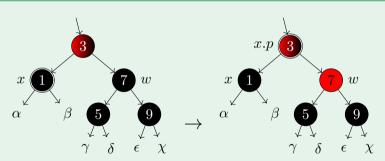
• Viene colorato di nero: ripristina i valori di bh senza violazioni

Cancellazione - Caso 1: x è nero, con fratello w rosso ($\Rightarrow x.p$ nero)



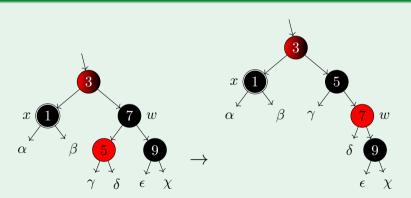
Scambio i colori di w e w.p; LEFTROTATE(x.p): x ha fratello w nero: \rightarrow casi 2,3,4

Cancellazione - Caso 2: x è nero, con fratello w nero, e nipoti entrambi neri (n.b. x.p può essere nero o rosso)



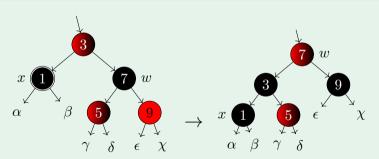
Coloro w di rosso (i.e. rimuovo un nero dal sottoalbero) e richiamo RIPARARBCANCELLA su x.p

Cancellazione - Caso 3: x è nero, con fratello w nero, e nipote dx nero (n.b. x.p può essere nero o rosso)



Scambio di colore w e w.left, RIGHTROTATE su $w o \mathsf{Caso}$ 4

Cancellazione - Caso 4: x è nero, con fratello w nero, e nipote dx rosso (n.b. x.p e w.left possono essere neri o rossi)



w prende il colore di w.p, w.right diventa nero. Invoco LEFTROTATE su w.p

Operazioni su alberi RB

RIPARARBCANCELLA – Analisi di complessità

- I casi 0,1,3 e 4 di RIPARARBCANCELLA effettuano un numero costante di rotazioni e scambi di colore o sono $\mathcal{O}(k)$
- ullet L'unica chiamata ricorsiva è quella che avviene nel caso 2 sul padre di x
 - Nel caso il padre ricada nei casi 0,1,3 o 4, la nuova chiamata termina in $\mathcal{O}(k)$
 - In caso contrario, viene ri-invocata RIPARARBCANCELLA
- Ad ogni chiamata ricorsiva si risale di un livello verso la radice \to al massimo effettuiamo $\mathcal{O}(\log(n))$ chiamate
- La procedura complessiva di cancellazione da alberi RB è quindi $\mathcal{O}(\log(n))$ come tutte le altre azioni