CSE 321 Discrete Structures

January 25, 2010

Lecture 09: Structural Induction

Structural Induction

- Show P holds for all basis elements of S.
- Show that P holds for elements used to construct a new element of S, then P holds for the new elements.

Prove all elements of S are divisible by 3

- Basis: $6 \in S$; $15 \in S$;
- Recursive: if $x, y \in S$, then $x + y \in S$;

Structural Induction

Let $\Sigma=\{a,b\}$; think of a = “(“ and b = “)”

Recall our toy language $L \subseteq \Sigma^*$ of well-formed parentheses:

- $\varepsilon \in L$
- If x, y, z $\in$ L then xaybz $\in$ L

Prove that every string in L has the same number of a’s and of b’s

Structural Induction

For every w ∈ L denote A(w) and B(w) the number of a's and of b's in w.

We prove that A(w) = B(w) by structural induction on w

- If w = ε, then A(w) = B(w) = 0
- If x, y, z ∈ L and w = xaybz, then:
  - The inductive hypothesis is:
    A(x) = B(x) and A(y) = B(y) and A(z) = B(z)
  - A(w) = A(x) + A(y) + A(z) + 1
  - B(w) = B(x) + B(y) + B(z) + 1
  - Hence A(w) = B(w)

Recursive Functions on Trees

- N(T) - number of vertices of T
- $N(\varepsilon) = 0$; $N(\bullet) = 1$
- $N(\bullet, T_1, T_2) = 1 + N(T_1) + N(T_2)$

- Ht(T) – height of T
- $Ht(\varepsilon) = 0$; $Ht(\bullet) = 1$
- $Ht(\bullet, T_1, T_2) = 1 + \max(Ht(T_1), Ht(T_2))$

NOTE: Height definition differs from the text
Base case $H(\bullet) = 0$ used in text

Structural Induction

- Prove that for every non-empty binary tree T, the following holds:

$$N(T) \leq 2^{Ht(T)} - 1$$

Structural Induction

Claim: $N(T) \le 2^{Ht(T)} - 1$

- If T =• then N(•) = 1, Ht(•) = 1; claim holds

- If T = (•, T_1, T_2), let $x = Ht(T_1)$, $y = Ht(T_2)$
  - By induction: $N(T_1) \le 2^x - 1$ and $N(T_2) \le 2^y - 1$
  - $N(T) = 1 + N(T_1) + N(T_2)$ and $H(T) = 1 + \max(x, y)$

$$N(T) \le 1 + 2^x - 1 + 2^y - 1 = 2^x + 2^y - 1$$
$$\le 2^{Ht(T)-1} + 2^{Ht(T)-1} - 1 = 2^{Ht(T)} - 1$$

The Importance of the Height

- $N(T) \le 2^{Ht(T)} - 1$ implies the following important property of binary trees:

$$Ht(T) \ge \log(N(T) + 1)$$

- What about the upper bound: ?

$$Ht(T) \le \ ??$$

- For most algorithms we want Ht(T) "small":

$Ht(T) \approx \log(N(T))$ is GREAT; $Ht(T) \approx N(T)$ is BAD

Fully balanced binary trees

- ε is a FBBT.
- if T_1 and T_2 are FBBTs, with $Ht(T_1) = Ht(T_2)$, then $(\bullet, T_1, T_2)$ is a FBBT.

- Prove at home that in a FBBT:

$$N(T) = 2^{Ht(T)} - 1$$

- This is nice, BUT: in practice can't keep trees fully balanced…

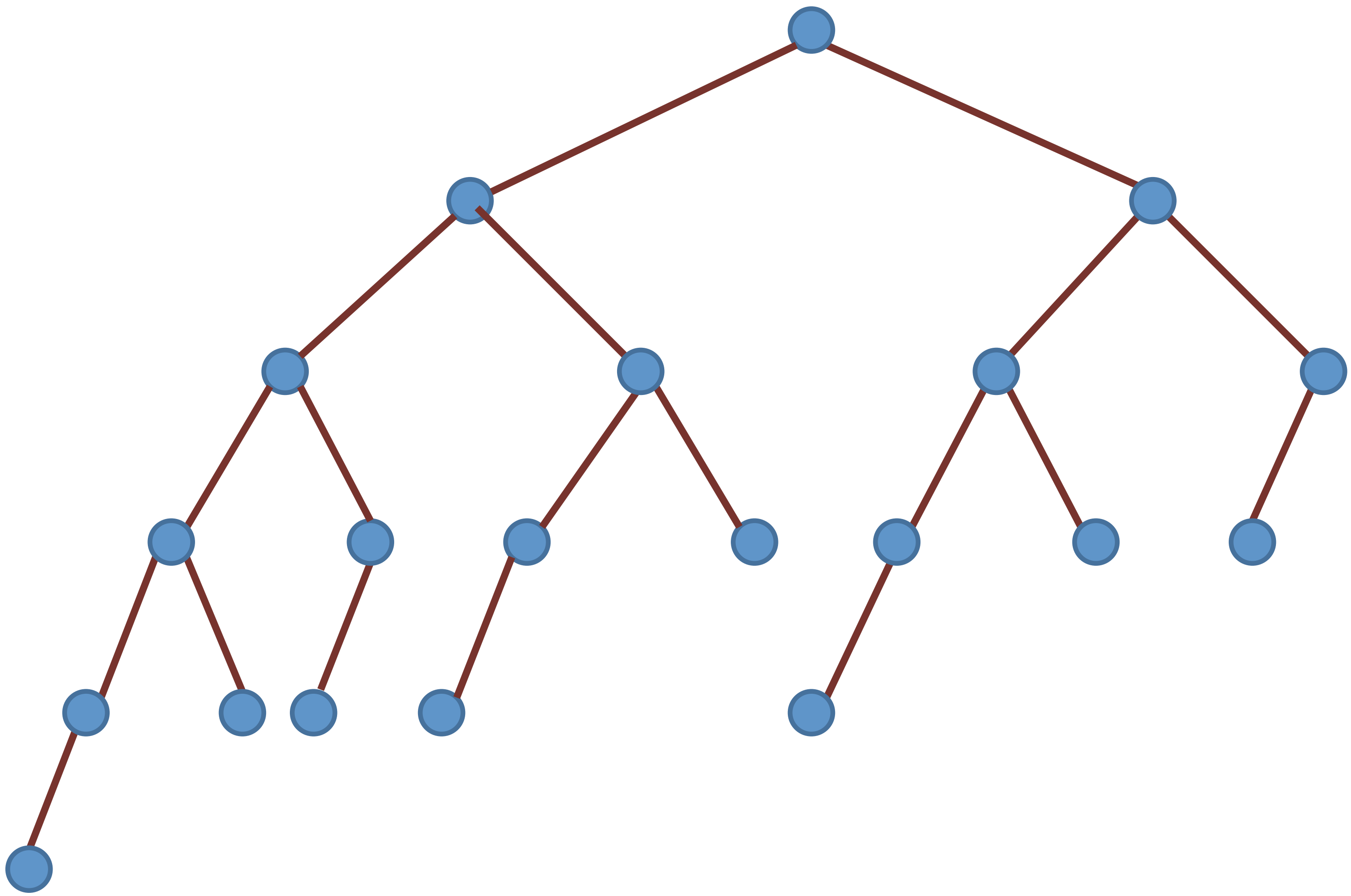
Almost balanced trees

Recursive definition:

- ε is a ABT.
- if T_1 and T_2 are ABTs with
  $Ht(T_1) - 1 \leq Ht(T_2) \leq Ht(T_1)+1$
  then $T = (\bullet, T_1, T_2)$ is a ABT.

Let $k_1 = Ht(T_1)$, $k_2 = Ht(T_2)$. Three cases:
$k_1 = k_2 + 1$ or $k_1 = k_2$ or $k_2 = k_1 + 1$

Is this Tree Almost Balanced?



Almost balanced trees

- So an "almost balanced tree" T can be quite imbalanced !
- Do we actually have Ht(T) $\approx$ log(N(T)) ?

Almost Balanced Binary Trees

Let f_k be the following sequence:

$$g_0 = 0, \quad g_1 = 0, \quad g_k = 1 + g_{k-1} + g_{k-2}$$

(we will compute later the sequence g_k)

Let T be an almost balanced tree. Prove the following:

If $n = N(T)$ and $k = Ht(T)$ then $n \geq g_k$

Structural induction on T:

If n = N(T) and k = Ht(T) then $n \geq g_k$

If T =ε, then n = 0, k = 0, and $0 \geq g_0$

If T = (•, T_1, T_2); let $n_i = N(T_i)$, $k_i = Ht(T_i)$, for i=1,2;
By induction we know $n_i \geq g_{ki}$

- Case 1: $k_1 = k_2 + 1$. Then $k = k_1 + 1$ and:
  $n = 1 + n_1 + n_2 \geq 1 + g_{k1} + g_{k2} = 1 + g_{k-1} + g_{k-2} = g_k$
- Case 2: $k_1 = k_2$. Then $k = k_1 + 1$ and:
  $n = 1 + n_1 + n_2 \geq 1 + g_{k1} + g_{k2} = 1 + g_{k-1} + g_{k-1} \geq g_k$
- Case 3: $k_2 = k_1 + 1$. Then $k = k_2 + 1$ and:
  $n = 1 + n_1 + n_2 \geq 1 + g_{k1} + g_{k2} = 1 + g_{k-2} + g_{k-1} = g_k$

Where did we use g_1 ? In Case 2: $1 + 2g_{k-1} \geq g_k$; $g_1 \leq 1+2g_0 = 1$

Solving Recurrences

$$f_0 = 1, \quad f_1 = 1, \quad f_k = f_{k-1} + f_{k-2}$$

Characteristic equation: $x^2 - x - 1 = 0$

Roots: $x_{1,2} = (1 \pm \sqrt{5})/2$

$$f_k = A\, x_1^k + B\, x_2^k$$

Solve A, B from initial conditions: $f_0=1$, $f_1=2$

$$f_k = 1/\sqrt{5}\,[(1 + \sqrt{5})/2]^k - 1/\sqrt{5}\,[(1 - \sqrt{5})/2]^k$$

Dominant term

Negative, but tiny

Solving Recurrences

Solve:

$$g_0 = 0, \qquad g_1 = 0, \qquad g_k = 1 + g_{k-1} + g_{k-2}$$

Let $f_k = g_k + 1$. Then:

$$f_0 = 1, \qquad f_1 = 1, \qquad f_k = f_{k-1} + f_{k-2}$$

We have solved f_k already; thus $g_k = f_k - 1$

Almost Balanced Binary Trees

What is its height ?

If $n = N(T)$ and $k = Ht(T)$ then
$n \geq g_k = f_k - 1 \approx 1/\sqrt{5}\, [(1 + \sqrt{5})/2]^k - 1$

$\log(n+1) \geq k \log [(1 + \sqrt{5})/2] - \log \sqrt{5}$

$k \lesssim \log(n+1) \;/\; \log [(1 + \sqrt{5})/2]$