6.851 Advanced Data Structures (Spring'10)

Prof. Erik Demaine Dr. André Schulz TA: Aleksandar Zlateski

Problem 4 Sample Solutions

Analyzing Partition Trees. Let r be the upper bound of $2(c\sqrt{2})^{\frac{1}{\epsilon}}$. Out of r partitions at most $c\sqrt{r}$ of them will cross a line. Therefore we get:

$$T(n) \leq O(r) + c\sqrt{r}T(2n/r)$$

Solving the equation using the Master Theorem, we get $T(n) \leq \Theta(n^{\frac{1}{2}+\epsilon})$.

Speed Up With LCP Array. Finding the LCP for $SA[i]$ and $SA[i+1]$ breaks down into three cases. First case when $SA[i]$ and $SA[i+1]$ are both from $T_0 \cup T_1$. Then we can calculate their LCP by looking up their locations in the SA for $\tilde{T}$, and getting the RMQ in the LCP array for $\tilde{T}$. When $SA[i]$ and $SA[i+1]$ are both from $T_0 \cup T_2$ we compare the first character of both strings. If they are different, we know that the LCP is 0, otherwise, the LCP is 1 plus the LCP of the two strings without their first characters. Finally if $SA[i]$ and $SA[i+1]$ are both from $T_1 \cup T_2$ we comparse the first two characters of both strings. If they differ in some way, we can set the LCP to either 0 or 1, depending on how many of the characters match. Otherwise, we proceed as in the previous case.

The first two cases take O(1) plus the time for the recursion. For the third case, when comparing equal characters L_s is increased by one. The total number of comparisons if equal chars can't be larger than $|P|$. The number of other comparisons in the last case is $O(1)$ per level of recursion. Hence the running time of $O(|P| + \log |T|)$

6.851 Advanced Data Structures

Spring 2010