Building suffix trees in $O(|T| + |\Sigma| \lg |\Sigma|)$ [Farach-FOCS 1997]

6.897
L 11.3
Apr. 20, 2003

- initial goal: compute sorted order of suffixes & lcp between adjacent suffixes

① sort Σ in $O(|\Sigma| \lg |\Sigma|)$ — can prove matching lower bound

② replace each letter in T by its rank in Σ
$\Rightarrow T'$ is a string of n letters from $\{1, \ldots, n\}$ [$n = |T|$]

③ coalesce adjacent letters in T'
$\Rightarrow T''$ is a string of $n/2$ letters from $\{1, \ldots, n^2\}$

④ sort those letters
& replace each letter in T'' by its rank in $\{1, \ldots, n^2\}$
$\Rightarrow T'''$ is a string of $n/2$ letters from $\{1, \ldots, n/2\}$.

⑤ recurse on T''' to get - sorted order of its suffixes
- lcp's of adjacent suffixes.

⑥ reversing $T''' \to T'' \to T' \to T$,
we get sorted order of even suffixes $(0, 2, 4, \ldots)$

⑦ get lcp's of adjacent even suffixes
by trying to extend recursive lcp's by 1.

(can't afford to recurse on odd suffixes)

- but each odd suffix = 1 char + even suffix

⑧ so get sorted order of odd suffixes by
stable sorting even suffixes by new 1st char.
(last step in radix sort)

⑨ lcp of adjacent pair of odd suffixes
= { range min. query of even-suffix lcp array
in range between corresp. even suffixes
+1 if added letters match first
0 else