# Constructing suffix tree in $O(|T| + sort(\Sigma))$

[Kärkkäinen & Sanders – ICALP 2003]

based on [Farach – FOCS 1997; Farach-Colton, Ferragina, Muthukrishnan → JACM 2000]



① sort $\Sigma$ – initially, in $O(sort(\Sigma))$ time
e.g. $O(|\Sigma| \lg |\Sigma|)$
– later, radix sort will suffice

② replace each letter of $T$ by its rank in $\Sigma$
(preserves order/lcp of suffixes)

③ form $T_0 = \langle (T[3i], T[3i+1], T[3i+2]) \text{ for } i = 0, 1, 2, \ldots \rangle$
$T_1 = \langle (T[3i+1], T[3i+2], T[3i+3]) \ldots \rangle$
$T_2 = \langle (T[3i+2], T[3i+3], T[3i+4]) \ldots \rangle$
$\Rightarrow$ suffixes$(T) \approx$ suffixes$(T_0) \cup$ suffixes$(T_1) \cup$ suffixes$(T_2)$

④ recurse on $\langle T_0, T_1 \rangle \Rightarrow$ size $= 2/3 \cdot n$
→ sorted order & lcps of suffixes$(T_0, T_1)$
$\Rightarrow$ suffixes$(T_0) \cup$ suffixes$(T_1)$

⑤ radix sort suffixes$(T_2)$
by writing $T_2[i:] = T[3i+2:] = (T[3i+2], \underbrace{T[3(i+1):]}_{T_0[i+1:]})$
– try to extend lcp's by 1 letter

⑥ merge suffixes$(T_0) \cup$ suffixes$(T_1)$ & suffixes$(T_2)$ via:

– $T_0[i:]$ vs. $T_2[j:]$
$= T[3i:]$ vs. $T[3j+2:]$
$= (T[3i], \underbrace{T[3i+1:]}_{=T_1[i:]})$ vs. $(T[3j+2], \underbrace{T[3(j+1)]}_{=T_0[j+1:]})$

– $T_1[i:]$ vs. $T_2[j:]$
$= T[3i+1:]$ vs. $T[3j+2:]$
$= (T[3i+1], T[3i+2], \underbrace{T[3(i+1):]}_{=T_0[i+1:]})$ vs. $(T[3j+2], T[3(j+1)], \underbrace{T[3(j+1)+1:]}_{=T_1[j+1:]})$

– try to extend lcp's by 1 or 2 letters

$\Rightarrow T(n) = T(2/3 \cdot n) + O(n) = O(n)$ $\quad (n = |T|)$