6.897
L18.2
Apr. 12, 2005

Suffix arrays

- sort suffixes of T
- just store their indices

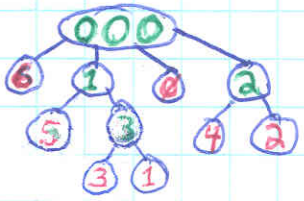
6	$	0
5	a $	1
3	a n a $	3
1	a n a n a $	0
0	b a n a n a $	0
4	n a $	2
2	n a n a $	

bananas$
0123456

- searchable in $O(|P| + \lg |T|)$ [as opposed to $O(|P|)$ for suffix tree]:
  - binary search
  - check letters only after what already matches in both L & R

Equivalence to suffix trees:

- in-order traversal of leaves in suffix tree → suffix array
- lcp array = |lcp|'s between consec. suffixes in sorted order
- Cartesian tree of lcp array (placing all min's at root)
  → internal nodes of suffix tree
  - suffixes fit in between as leaves
- lcp's give "letter depth" ⇒ length of edges labels
  ⇒ edge labels



- lcp array computable in $O(|T|)$ time [Kasai et al. - CPM 2001] or directly in suffix-array construction (below)