6.897
L 18.1
Apr. 12, 2005

String matching:

given text T & pattern P — strings over alphabet Σ
find some/all occurrences of P in T (as substring)

- one-shot: $O(|T|)$ time [Knuth, Morris, Pratt – SICOMP 1977;
  Boyer, Moore – CACM 1977;
  Karp, Rabin – IBM JRD 1987]

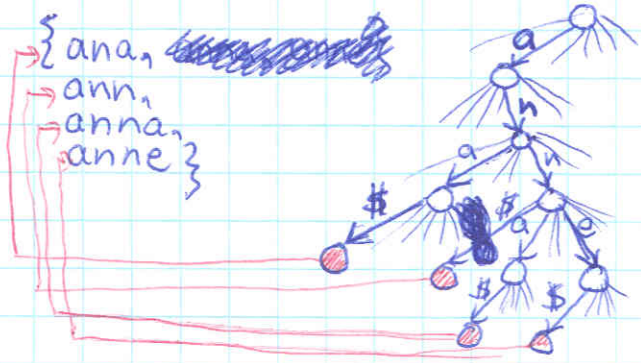
- preprocess $T \Rightarrow$ $O(|T|)$ space (words — later: $O(|T|)$ bits)
  $O(|T| + \text{sort}(\Sigma))$ preprocessing
  $O(|P|)$ query

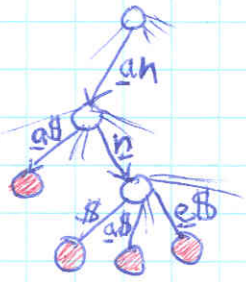
  ↳ SUFFIX TREE

Trie = tree with children branches labeled with letters from Σ
(not necessarily all present)

$\Leftrightarrow$ set of strings (root-to-leaf paths)
– terminate strings with \$ to distinguish prefixes

e.g. {ana, ann, anna, anne}



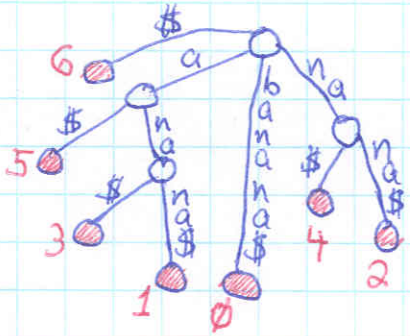


COMPRESSED

Compressed trie: store nonbranching paths with first letter in path

Suffix tree (trie): compressed trie of all $|T|+1$ suffixes of $T\$$

e.g. banana\$
0123456



– $|T|+1$ leaves
– store edge labels as 2 indices into text $\Rightarrow$ $O(|T|)$ space (words)
(or just store length & check later (& first letter))