Compact suffix arrays/trees

6.897
L21.5
Apr. 28. 2005

- Grossi & Vitter [STOC 2000]:

  $(\frac{1}{\varepsilon} + O(1))\,|T| \lg |\Sigma|$ bits

  $O(\frac{|P|}{\lg_{|\Sigma|} |T|} + \lg^{\varepsilon}_{|\Sigma|} |T| \cdot |\text{output}|)$ query

- Ferragina & Manzini [FOCS 2000]:

  $O(H_k(T) \cdot |T| + \frac{|T|}{\lg |T|} |\Sigma| \lg |\Sigma|) + o(|T|)$ bits $\forall$ fixed k

  → # bits/char. in code depending on last k letters

  kth-order empirical entropy:

  $$\sum_{|w|=k} \Pr[w \text{ occurring}] \cdot H_0(\text{string of successor chars of } w)$$

  ($\Pr[w \text{ occurring}]$ = # occurrences / $|T|$)

  $O(|P| + \lg^{\varepsilon} |T| \cdot |\text{output}|)$ query

- Sadakane [J. Alg. 2003]:

  $O(H_0(T) \cdot |T| + |\Sigma| \lg |\Sigma|) + o(|T|)$ bits

  $O(|P| \lg |T| + \lg^{\varepsilon} |T| \cdot |\text{output}|)$ query

- low-space construction: [Hon, Sadakane, Sung – FOCS 2003]

  $O(|T| \lg |\Sigma|)$ bits working space

  $O(|T|)$ time for suffix array

  $O(|T| \lg^{\varepsilon} |T|)$ time for suffix tree

- more suffix-tree like [Hon & Sadakane – CPM 2002]

  e.g. find maximal common substrings