O(n)-space suffix array

6.897
L22.3
Apr. 28, 2005

- store only $1/\varepsilon + 1$ levels of recursion:
  $0, \varepsilon\ell, 2\varepsilon\ell, \ldots, \ell = \lg\lg n$
- representing $SA_{k\varepsilon\ell}$ by $SA_{(k+1)\varepsilon\ell}$:
  - similar but "even" now means "in $SA_{(k+1)\varepsilon\ell}$" (i.e., even $\varepsilon\ell$ times)
  - bitvector stores which of $n_{k\varepsilon\ell}$ ($n_{(k+1)\varepsilon\ell}$) suffixes of $SA_{k\varepsilon\ell}$ are represented in $SA_{(k+1)\varepsilon\ell}$ (1 = yes)
  - rank_1 structure on this bit vector
  - $\text{succ}_{k\varepsilon\ell}(i) = j$ where $SA_{k\varepsilon\ell}[j] = SA_{k\varepsilon\ell}[i] + 1$
  - to compute $SA_{k\varepsilon\ell}[i]$:
    - follow succ repeatedly until j "in" $SA_{(k+1)\varepsilon\ell}$ using bit vector
    - recurse on rank of j in $SA_{(k+1)\varepsilon\ell}$, mult. by $2^{\varepsilon\ell}$
    - subtract by # calls to succ.
- search cost: walk up to $2^{\varepsilon\ell} = \lg^{\varepsilon} n$ per level
  $\Rightarrow O(\lg^{\varepsilon} n \lg\lg n) = O(\lg^{\varepsilon'} n)$
  OPEN: improvable with $O(n)$ space?
- space: $n_{k\varepsilon\ell} + O(n_{k\varepsilon\ell} \frac{\lg\lg n_{k\varepsilon\ell}}{\lg n_{k\varepsilon\ell}})$ for bitvector & rank
  $\Rightarrow \le n + O(n \frac{\lg\lg n}{\lg n})$ total

  $n + 2n_{k\varepsilon\ell}$ for succ ~ essentially even-succ + odd-succ merged + some optimization
  $+ O(n_{k\varepsilon\ell}/\lg\lg n)$

  only $n + O(n_0/\lg\lg n)$ for succ @ $k=0$ ~ more optimization

  $\Rightarrow (\frac{1}{\varepsilon}+2)n + O(n/\lg\lg n)$ total
- improvement: use static dictionary instead of bitvector
  [Brodnik & Munro – SICOMP 1999; Pagh – SICOMP 2001]
  - space: $\lg \binom{n_{k\varepsilon\ell}}{n_{(k+1)\varepsilon\ell}} + O(n_{k\varepsilon\ell} \frac{(\lg\lg n_{k\varepsilon\ell})^2}{\lg n_{k\varepsilon\ell}})$

    $\approx n_{(k+1)\varepsilon\ell} \lg \frac{n_{k\varepsilon\ell}}{n_{(k+1)\varepsilon\ell}} \underbrace{\frac{2^{\varepsilon\ell}}{\varepsilon\ell}}$

    $= O(n/\lg\lg n)$ $\Rightarrow (\frac{1}{\varepsilon}+1)n + O(n/\lg\lg n)$ total.