

Enhanced and Extended Suffix Arrays

Adrian Regenfuß

Technische Universität München

Sommersemester 2020

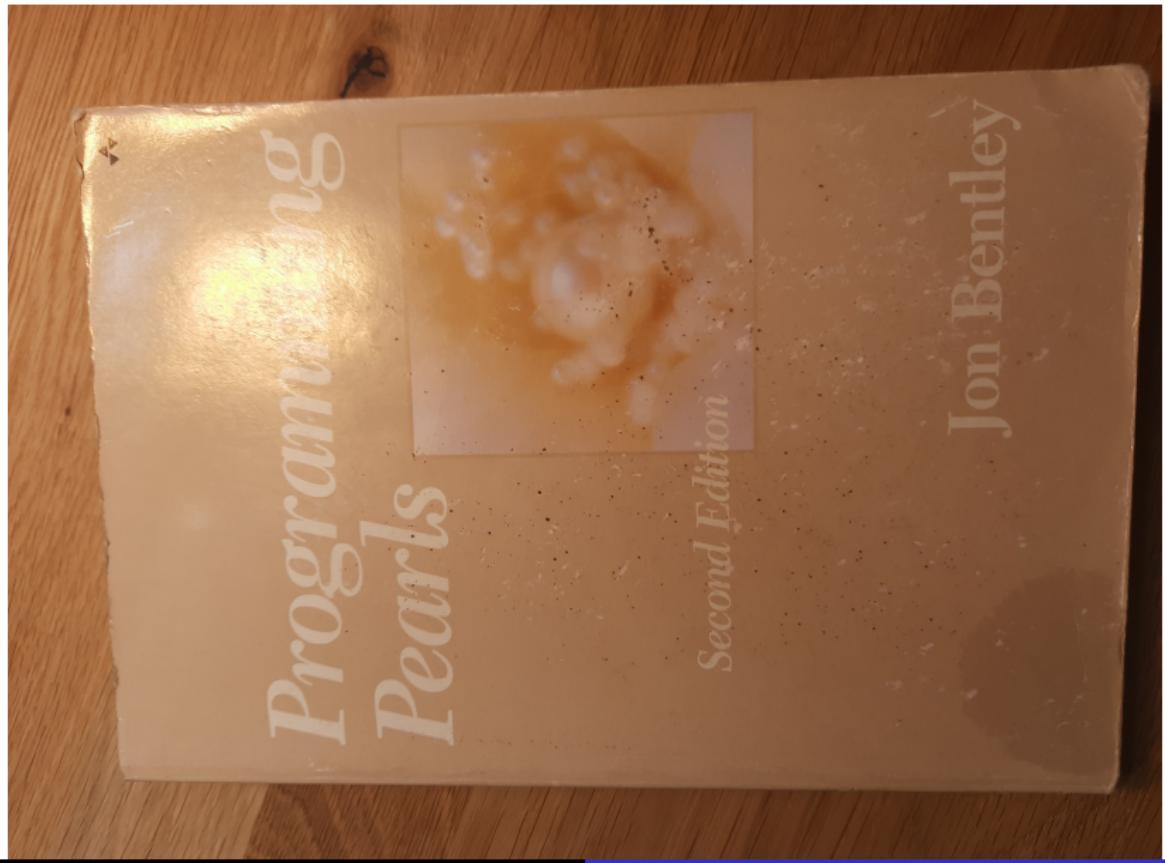
Übersicht

- 1 A Personal Anecdote
- 2 Structure
- 3 Data Structures
- 4 Terminology
- 5 Algorithms
- 6 Comparison with the Suffix Tree

4 years ago

- Me in Nepal
- No computer with me, only pen, paper and Programming Pearls

My old, damaged copy of Programming Pearls



Thinking a lot about Algorithms

2017-04-16: 01
int lin-find (B[], n, k)

if $n = 0$ & $k \geq 0$
 ret 0

else if $k[n-1] \leq k$ $i = (B[n-1] - B[0]) / (k - B[0]) + 1$
 ret i

$\ell = 0$

$v = n - 1$

loop

$m = \text{ceil}((B[v] - B[0]) / (k - B[\ell])) + \ell$

if $B[m] \geq k \& \& B[m-1] \leq k$

 ret m

 else if $B[m] < k$

$\ell = m$

 else $v = m - 1$

$\ell = v$

$m = \text{ceil}((B[v] - B[0]) / (k - B[\ell])) + \ell$

 if $B[m] \geq k \& \& B[m-1] \leq k$

 ret m

 else if $B[m] < k$

$\ell = m$

 else

$v = m - 1$

 ret v

int lin-find (B[], n, k) kind lin-find (B[], n, k)
for $i = 0$; $i < n$ & $B[i] < k$; $i++$
 ret i

Problems back then

- Thinking about searching, sorting, collision detection, string matching
- Specifically: What is the longest repeated substring of a string?
- Or: what is the longest supermaximal repeat of a string?
- Devising very complex algorithms

Discovering Suffix Arrays

- On the flight back reading the 15th chapter of Programming Pearls
- Describes this exact problem, solved using suffix arrays

Übersicht

- 1 A Personal Anecdote
- 2 Structure
- 3 Data Structures
- 4 Terminology
- 5 Algorithms
- 6 Comparison with the Suffix Tree

Structure

- Description of data structures
- Clarification of terminology
- Description of algorithms
- Discussion of advantages/disadvantages

Übersicht

- 1 A Personal Anecdote
- 2 Structure
- 3 Data Structures
- 4 Terminology
- 5 Algorithms
- 6 Comparison with the Suffix Tree

Suffix Array suftab

LCP Array | cptab

Burrows-Wheeler Transformation Array bwttab

Inverse Suffix Array suftab^{-1}

LCP-Intervals

LCP-Interval Trees

Übersicht

- 1 A Personal Anecdote
- 2 Structure
- 3 Data Structures
- 4 Terminology
- 5 Algorithms
- 6 Comparison with the Suffix Tree

Enhanced Suffix Array vs. Extended Suffix Array

Übersicht

- 1 A Personal Anecdote
- 2 Structure
- 3 Data Structures
- 4 Terminology
- 5 Algorithms
- 6 Comparison with the Suffix Tree

Exact String Matching (Substring Search)

Finding Maximal Repeats

Finding Supermaximal Repeats

Finding Longest Common Substrings

Finding Maximum Unique Matches

Übersicht

- 1 A Personal Anecdote
- 2 Structure
- 3 Data Structures
- 4 Terminology
- 5 Algorithms
- 6 Comparison with the Suffix Tree

Of the Suffix Array

Of the LCP Array

Of the BWT Array

Übersicht

- 1 A Personal Anecdote
- 2 Structure
- 3 Data Structures
- 4 Terminology
- 5 Algorithms
- 6 Comparison with the Suffix Tree