


B w T

2025-11-17



BURROWS - WHEELER TRANSFORMAT

egyfajta preprocessing

INPUT: string $\in \{A-Z\}^*$

OUTPUT: string $\in \{A-Z, \$\}^*$

BWT: 1.) INPUT + \$

BANANA \mapsto BANANA\$

2) ÖSSZES CIKLIKUS ELTOCTAT LEBRÁKTJUK:

BANANA\$

ANANA\$B

NANA\$BA

\vdots

\$BANANA

3) Ezeket ábécarendbe rendez

$\$ < A < B < \dots < Z$

\$BANANA
A\$BANAU
↓ ANA\$BAN
ANANA\$B
BANANA\$
NA\$BAN
NANA\$BA

h) OUTPUT := UTOLSÓ OSZLOP : ANUB\$AA

Mire jó

1) TÖMÖRÍTÉS EGYIK LÉPÉSE

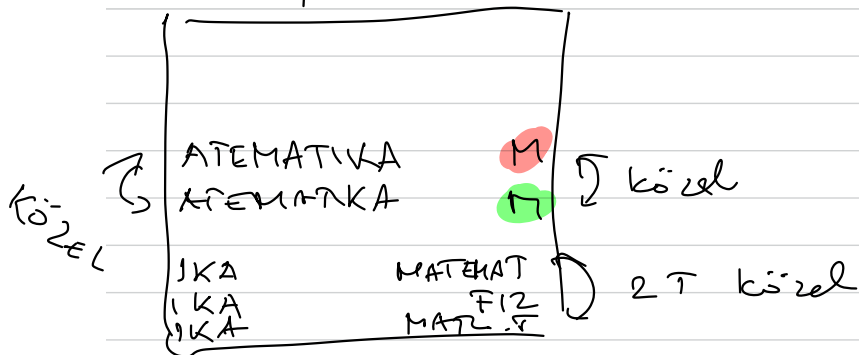
indoklás: szöveg

Csak olyan szöveg tömöríthető, ami valahogy „szabályos”

pl. ismétlődésel van!:

" M A T E M A T I K A M A T H M A T I K A . -
 . . . I K A ~ ✓

↓ BWT



utolsó azaz:

hosszú futamok

alacsony betűelő-

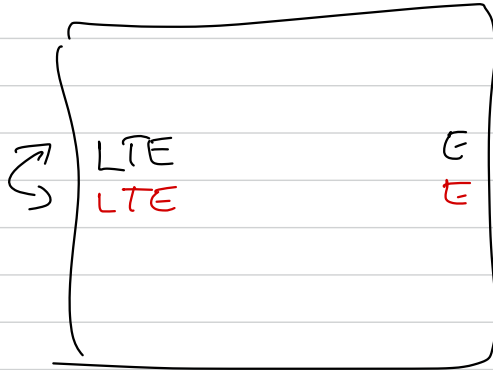
input → BWT → Last-to-front → SZÁM SZÖVEG :
HUFFMAN-KÓD

bzip2

2. két stringre hosszú közös prefix keresése

STRET

$S_1 + S_2$ konstans \rightarrow BWT



\rightarrow MATCH

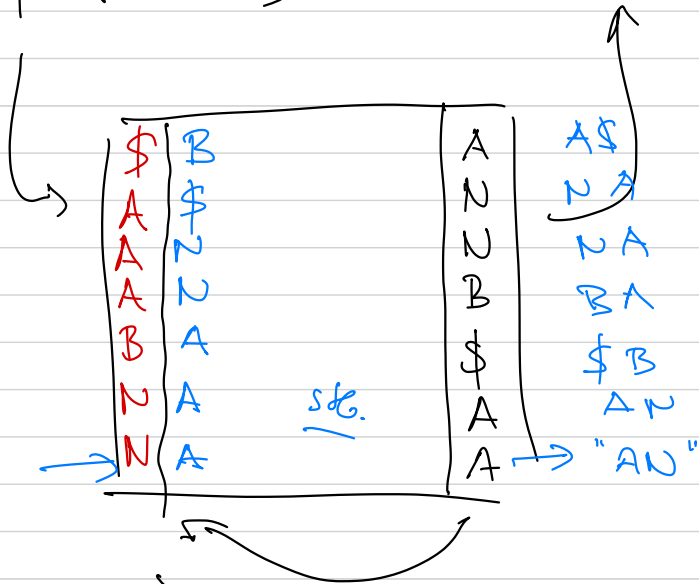
$ELTE = E LTE$

Eleg szomszédos pont

között keresni az egyenest.

$w \mapsto \text{BWT}(w)$ ✓ DE ПОСЛЕДН ЛЕВ ПРАВЕКОМ

$ANNB\$AA \xrightarrow{?}$



1. l'èpe:

1. onlop: ugyanaz, de ábrázol.

2.1 epis: MEGUANNAK

a 2 hosti řezel
→ 1-2. ořez

HATÉKONY(ABB) ALGO.

naiv algo: $n \rightarrow n \times n$ db whig $\xrightarrow{\text{SOU}}$ TÁBLÁZAT $\xrightarrow{n.}$ \rightarrow $n.$ \rightarrow $n.$

?

$O(n \cdot \log n)$ összehasonlítás

↓
hívás

$O(n)$

idő \leq ?

WORST CASE

$O(n^2 \cdot \log n)$ IDŐ

TÁBL.: $O(n^2)$

AMI NEHEZ: BWT $O(n)$ időben, $O(n)$ helyen

ETHELYETT: $O(n \cdot \log n)$
↓

TAR

- A táblázat helyett csak az adott sorokban látható eltolások kérdőpozícióit tartjuk.

6. \$BANANA
5. A\$BANAU
3. ANA\$BAN
7. ANANAS\$
0. BANANA\$
4. NA\$BAN
2. NANA\$BA

6 5 3 1 0 4 2

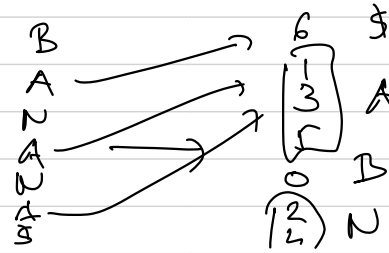
IDŐ

- Rekurzívan mindig duplázunk a táblázat kinézetét. $1 \rightarrow 1..2 \rightarrow 1..4$

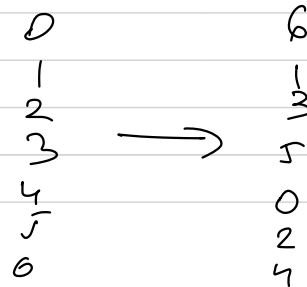
ALGO (A2L45) :

1. feladat sorbarendezés az 1. karakter alapján

BANANA\$ → DA -- 0.
 AN --- 1.
 \$ --- 6.



VEGES
 ↓
 ABECE



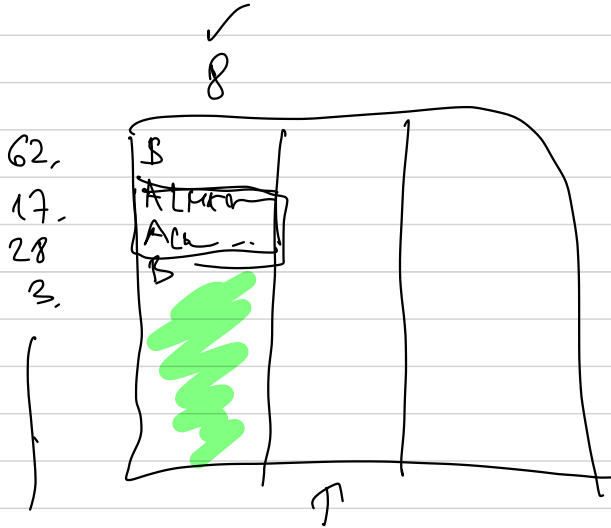
LIN SÖPRES:

$O(n)$

REKURZÍV LÉPÉS: ^{PL.} MÁR 8 ÖZLÖP RENDEZVE



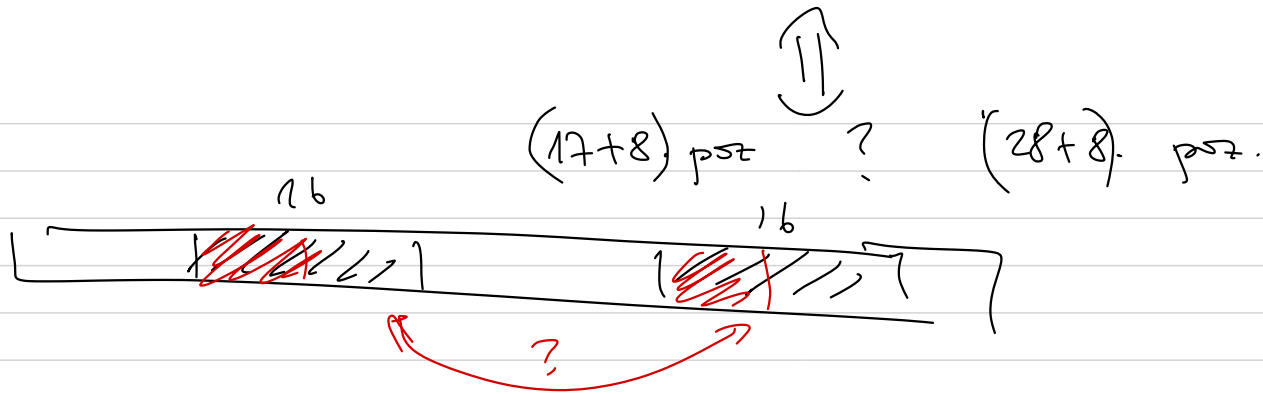
16 ÖZLÖP RENDEZVE



CSAK AKKOR SZÁMÍT, HA AZ ELŐŐ 8 ÖZLÖP
ALAPJÁN DÖNTETLEN

DÖNTETLENNEK: 17. poz $\begin{matrix} > \\ = \\ < \end{matrix}$ 28. poz

16 betű alapján



REKURZÍV LÉPÉS: $O(n)$

MEG NÉHA'NY KÉRDÉS, MIÉRT A BWT VÉGÉN (NOKON)

SUFFIX ARRAY RÖGÍT:

Q.) MI A LEGHOSSZABB PALINDROM EGZ KÖZBEN $O(n)$!

TRÜKK $w, w^R \rightarrow$ közös BWT
|
TÜKÖR

Q) ELŐFORDULÓ KÜL. RÉSZTRÉSEK SAJÁ + ITB.