


BURROWS - WHEELER - transzformáció

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

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

BANANA \mapsto

1. lépés: BANANA\$

2. lépés: összes ciklikus permut.

BANANA\$
ANANA\$B
NANA\$BA
ANA\$BAN
NA\$BANA
A\$BANAN
\$BANANA

sort (ábécésorrendbe)

3. lépés. $\$ < A < B < \dots < Z$

3. utal:

\$BANANA
A\$BANAN
ANAS\$BAN
ANANAS\$B
BANANAS\$
NAS\$BANA
NANAS\$BA

4. lépés (output): utolsó overlap ANNB\$AA

ALK. 1 : TÖMÖRÍTÉS ELSŐ LÉPÉSE .

input ... - MATEMATIKA ... - MATEMATIKA ...

... M ...

... -

MA

→ BWT

ATEMATIKA w A _ _

ATEMATIKAI _ _ _

AT _ _ _ A _ _

M
M
M

← sok egyforma betű

bzip2 : BWT + MOVE-TO-FRONT + HUFFMANN



szoftár: 0 1 2 25
 1 ↓ 1/2 ↓
 A B c 2

előpróba : S:
 0 1 2 3
 ↓ ↓ ↓ ↓
 S A B c _ _

MM _ _ M
↓ ↓ ↓ ↓ ↓
12 0 0 0 _ 0

THE

THE

SHE

THE

SHE

THE

AHEM

HE ?
HE ?
HE ?
HE ?
HE ?

T
T
A
S
T

tömör

vesztéséges

vesztéssegmentes

↕
?
↕

BWT

6
5
3
4
0.
4
2
↑

ROT.
MEUNTSÉGE

\$BANANA
A\$BANAN
ANAS\$BAN
ANANAS\$B
BANANAS\$
NAS\$BAN
NANAS\$BA



←

ANNUS\$AA

?

BANANA

← ----

← visszaépítjük a táblázatot

\$	B		A
A	\$		N
A	N		N
A	N		B
B	A		\$
N	A		A
N	A		A

A\$
AN
AN

lebecskolt,
B az
első

→



utolsó + első → ugyanaz, csak sorrendben
2 hosszú rész OK

utolsó + 2első \rightarrow 3kosn' rest \rightarrow 3orlop \rightarrow stb.

—
Először a naiv megközelítéssel a futási ideje / tárhelye?

$$\text{sort} \left(\begin{matrix} n \text{ db} \\ n \text{ kosn' chg} \end{matrix} \right) \quad O(n^2)$$

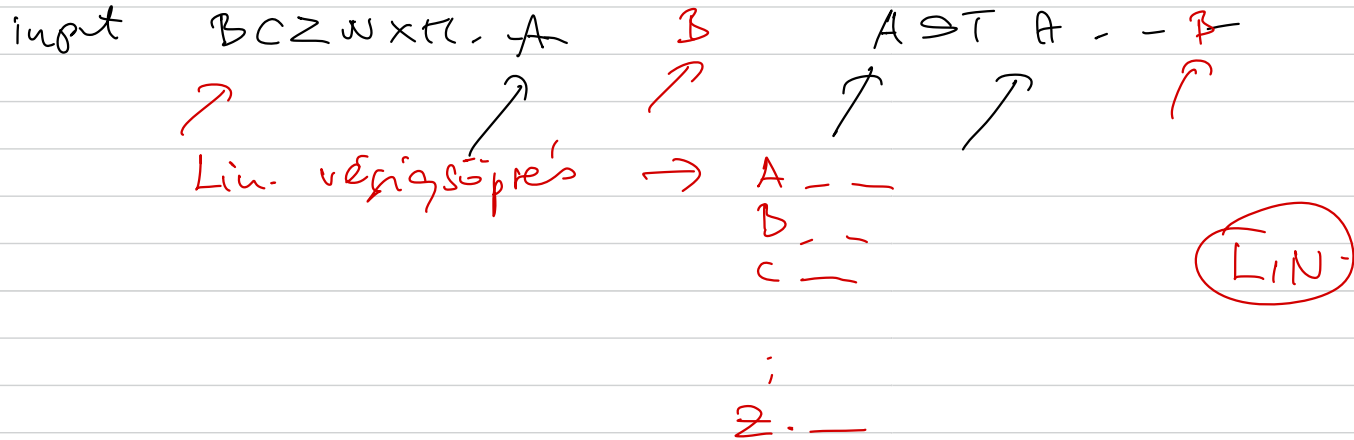
$$O(n \cdot \log n \cdot \text{összehasonlítás ideje})$$

—
CSÖKKENTENI KELL!

TÁRHELYHEZ: a táblázat helyett elég a rotáció mértékét minden sorhoz.

KISZÁMÍTÁSTÓZ : rekurzív, minden lépésben duplázza a kiszámolt orlopok számát.

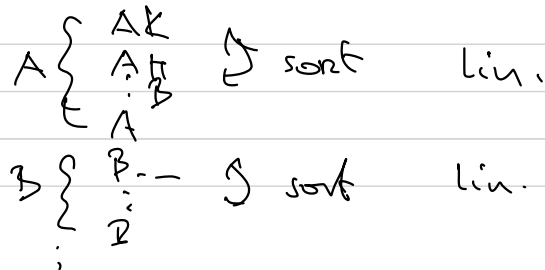
1. fázis: sorbarendezés az 1. karakter alapján



2. fázis: DETÜPÁROLK

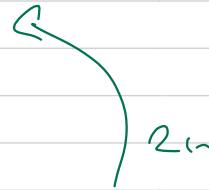
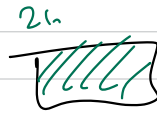
(csak kordöbetük fixálása után)

LIN



3. fázis: 4 hosni

4. fázis: 8 hosni



$$\frac{\vdots}{O(n \cdot \log n)}$$

BRUTÁL MÓDRA $\rightarrow O(n)$

ALK 2. : SZTRINGES KÉRDÉSEK GYORS MEGVÁLASZOLÁJA
ELŐKEZLETES (PREPROCESSÁLÁS UTÁN).

BWT ~ SUFFIX TÖMB \leftrightarrow SUFFIX FA
 \downarrow SZEMANTIKA
ADATMERK.

Q: ELŐFORDUL-E MINTA / HÁNTSOR CGTCGACT
 $\underbrace{\hspace{15em}}_n$ hossz \downarrow
 $O(n) \rightarrow O(n)$

Q: MI A LEGHOSSZABB PALINDROM BENNE?

SUFFIX-FA: SZÓTÁR A SZÓ SUFFIXEIBŐL + TÖMÖRÍTÉS

SUFFIXEK:

BANANA →

BANANA\$

ANANA\$

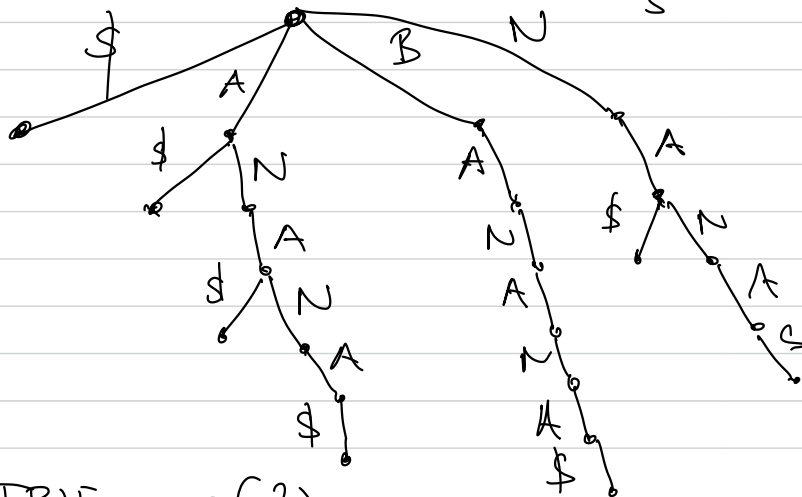
NANA\$

ANA\$

NA\$

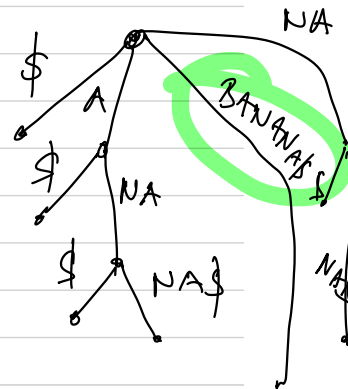
A\$

S



SUFFIX TRIE

$O(n^2)$



SUFFIX TREE

BETŰK pozíció
ÉLCÍMKÉK: BANANAS $\rightarrow [0,6]$

NANA $\rightarrow [2,5]$

$O(n)$ tárigény, TRÜKKÖS ALGO: $O(n \log n)$

+ NAGYA $O(n)$

leggyorsabb

HA MÁR MEGVAN:

Pl. keresés: NAB?

a gyökér fölől elindulva
keresés

PALINDROM KERESÉSE :

Szö : w

szó vissrafelé: w^R

ij-keuze test

$$\rightarrow \overbrace{w \# w}^{\downarrow}$$

BWT :

W

.. - GöRÖGÖK

$$\sim n$$
$$- \kappa \hat{\sigma}_0 \rho \hat{\sigma}_0 G, \dots$$

G	O	R	O	G
G	O	K	O	G

$$\frac{\omega}{\omega_R}$$

6

ALFA U
ALMA U