

# Suffiksitaulukko

Algoritmikerho 19.10.

# Leksikografinen järjestys

- Aakkosjärjestys
- \$ pienempi kuin muut

ananas\$

anastaa\$

anana\$

ananas\$

# Suffiksit, loppuosat

ananas\$ →

|   |          |
|---|----------|
| 0 | ananas\$ |
| 1 | nanas\$  |
| 2 | anas\$   |
| 3 | nas\$    |
| 4 | as\$     |
| 5 | s\$      |
| 6 | \$       |

# Suffiksit leksikografisessa järjestyksessä

ananas\$ →

|   |          |
|---|----------|
| 6 | \$       |
| 0 | ananas\$ |
| 2 | anas\$   |
| 4 | as\$     |
| 1 | nanas\$  |
| 3 | nas\$    |
| 5 | s\$      |

# Suffiksitaulukko (suffix array)

s = ananas\$

| i | SA[i] | s[SA[i]:] |
|---|-------|-----------|
| 0 | 6     | \$        |
| 1 | 0     | ananas\$  |
| 2 | 2     | anas\$    |
| 3 | 4     | as\$      |
| 4 | 1     | nanas\$   |
| 5 | 3     | nas\$     |
| 6 | 5     | s\$       |

# Kaksinkertaistusmenetelmä (prefix doubling)

| JL[i] | i | s[i:]    |
|-------|---|----------|
| 0     | 6 | \$       |
| 1     | 0 | abcaba\$ |
| 1     | 5 | a\$      |
| 1     | 3 | aba\$    |
| 2     | 1 | bcaba\$  |
| 2     | 4 | ba\$     |
| 3     | 2 | caba\$   |



| JL[i] | i | s[i:]    |
|-------|---|----------|
| 0     | 6 | \$       |
| 1     | 5 | a\$      |
| 2     | 0 | abcaba\$ |
| 2     | 3 | aba\$    |
| 3     | 4 | ba\$     |
| 4     | 1 | bcaba\$  |
| 5     | 2 | caba\$   |

# Kaksinkertaistusmenetelmä (prefix doubling)

abcaba\$  
aba\$

- "ca" vastaa suffiksia "caba\$", "a\$" itseään
- Järjestetään parien (JL[i], JL[i + 2]) mukaan, saadaan uudet järjestysluvut

| JL[i] | i |        | s[i:]    |
|-------|---|--------|----------|
| 0     | 6 | (0, 0) | \$       |
| 1     | 5 | (1, 0) | a\$      |
| 2     | 0 | (2, 5) | abcaba\$ |
| 2     | 3 | (2, 1) | aba\$    |
| 3     | 4 | (3, 0) | ba\$     |
| 4     | 1 | (4, 2) | bcaba\$  |
| 5     | 2 | (5, 2) | caba\$   |

# LCP-taulukko (longest common prefix)

- Pisin yhteinen alkuosa suffiksien  $SA[i]$  ja  $SA[i-1]$  välillä

| i | SA[i] | LCP[i] | s[SA[i]:]        |
|---|-------|--------|------------------|
| 0 | 6     | –      | \$               |
| 1 | 0     | 0      | <u>a</u> nanas\$ |
| 2 | 2     | 3      | <u>a</u> nas\$   |
| 3 | 4     | 1      | as\$             |
| 4 | 1     | 0      | <u>n</u> anas\$  |
| 5 | 3     | 2      | nas\$            |
| 6 | 5     | 0      | s\$              |

# LCP-taulukko (longest common prefix)

anas\$  
anas\$  
anas\$



anas\$  
...nas\$  
nas\$