

# Algoritmikerho 16.3.

Juha Harviainen

**Kumpi helpompi esittää ytimekkäästi?**

ababababababababababababababab

aabaaabbbbaabaaaaabbaba

## Kumpi helpompi esittää ytimekkäästi?

ababababababababababababab

*“toista ab 12 kertaa”*

aabaaaabbbbaabaaaaaabbaba

*“???”*

# Pakkausalgoritmit

- **Tavoite:** Esitä annettu data tiiviimmässä muodossa
- Kuvia, ohjelmistoja, merkkijonoja, ...
- Vie vähemmän tilaa koneella
- Vähemmän siirrettävää netissä



# Voiko aina pakata?

- **Onko yleispätevää pakkausalgoritmia, joka aina pienentää kokoa?**
- Ei :(
- Lyhyempien bittijonojen joukkio jo “täynnä”, ei voi tunkea uusia
- 0, 1, 00, 01, 10, 11: jos 111 yritetään lyhentää, joku pitenee
- **Tilannekohtaisia**
- Datassa syytä olla rakennetta, satunnaisuutta vaikea tiivistää
- Rakenteesta riippuen eri algoritmit voivat toimia paremmin

# Pakkausalgoritmeja

- Käytännössä monet käytetystä monimutkaisia
- Joitakin yksinkertaisempia esimerkkejä

# Monta lukua samoihin tavuihin

- Tavu sisältää kahdeksan bittiä
- Leikitään että on kaksi nelibittistä symbolia/lukua

$$\begin{array}{r} 11 \quad 5 \\ 1011 \quad 0101 \\ 10110101 \\ (1011 \ll 4) + 0101 \\ (10110101 \gg 4) = 1011, \quad (10110101 \bmod 16) = 0101 \end{array}$$



- Run-length encoding eli jakson pituuden koodaus
- Korvataan peräkkäiset samat merkit niiden määrällä ja yhdellä merkillä
  - AAAABBBCCD = 4A3B2C1D
- Ongelmallinen jos tiheästi vaihtuvia merkkejä
  - ABCDBAC = 1A1B1C1D1B1A1C

# Lempel–Ziv (1977)

- Haetaan toistuvia pätkiä
- Jos pätkä löytyy, viitataan taaksepäin
- (d, l, c): etäisyys osuman alkuun, osuman pituus, sitä seuraava merkki

1	2	3	4	5	6	7	8	9	10	11
A	B	C	D	B	C	A	B	C	A	C

--	--	--	--	--	--	--	--	--	--	--

# Lempel–Ziv (1977)

- Haetaan toistuvia pätkiä
- Jos pätkä löytyy, viitataan taaksepäin
- (d, l, c): etäisyys osuman alkuun, osuman pituus, sitä seuraava merkki

1	2	3	4	5	6	7	8	9	10	11
A	B	C	D	B	C	A	B	C	A	C

$\emptyset, \emptyset, A$	$\emptyset, \emptyset, B$	$\emptyset, \emptyset, C$	$\emptyset, \emptyset, D$							
---------------------------	---------------------------	---------------------------	---------------------------	--	--	--	--	--	--	--

# Lempel–Ziv (1977)

- Haetaan toistuvia pätkiä
- Jos pätkä löytyy, viitataan taaksepäin
- (d, l, c): etäisyys osuman alkuun, osuman pituus, sitä seuraava merkki

1	2	3	4	5	6	7	8	9	10	11
A	B	C	D	B	C	A	B	C	A	C

$\emptyset, \emptyset, A$	$\emptyset, \emptyset, B$	$\emptyset, \emptyset, C$	$\emptyset, \emptyset, D$	3, 2, A						
---------------------------	---------------------------	---------------------------	---------------------------	---------	--	--	--	--	--	--

# Lempel–Ziv (1977)

- Haetaan toistuvia pätkiä
- Jos pätkä löytyy, viitataan taaksepäin
- (d, l, c): etäisyys osuman alkuun, osuman pituus, sitä seuraava merkki

1	2	3	4	5	6	7	8	9	10	11
A	B	C	D	B	C	A	B	C	A	C

$\emptyset, \emptyset, A$	$\emptyset, \emptyset, B$	$\emptyset, \emptyset, C$	$\emptyset, \emptyset, D$	3, 2, A	3, 3, C					
---------------------------	---------------------------	---------------------------	---------------------------	---------	---------	--	--	--	--	--

# Lempel–Ziv (1977)

- Pakkaus purettävissä vasemmalta oikealle

$\emptyset, \emptyset, A$	$\emptyset, \emptyset, B$	$\emptyset, \emptyset, C$	$\emptyset, \emptyset, D$	$3, 2, A$	$3, 3, C$					
---------------------------	---------------------------	---------------------------	---------------------------	-----------	-----------	--	--	--	--	--

1	2	3	4	5	6	7	8	9	10	11

# Lempel–Ziv (1977)

- Pakkaus purettävissä vasemmalta oikealle

0,0,A	0,0,B	0,0,C	0,0,D	3,2,A	3,3,C					
-------	-------	-------	-------	-------	-------	--	--	--	--	--

1	2	3	4	5	6	7	8	9	10	11
A										

# Lempel–Ziv (1977)

- Pakkaus purettävissä vasemmalta oikealle

0,0,A	0,0,B	0,0,C	0,0,D	3,2,A	3,3,C					
-------	-------	-------	-------	-------	-------	--	--	--	--	--

1	2	3	4	5	6	7	8	9	10	11
A	B	C	D							



# Lempel–Ziv (1977)

- Pakkaus purettävissä vasemmalta oikealle

0,0,A	0,0,B	0,0,C	0,0,D	3,2,A	3,3,C					
-------	-------	-------	-------	-------	-------	--	--	--	--	--

1	2	3	4	5	6	7	8	9	10	11
A	B	C	D	B	C	A				

# Lempel–Ziv (1977)

- Pakkaus purettävissä vasemmalta oikealle

0,0,A	0,0,B	0,0,C	0,0,D	3,2,A	3,3,C					
-------	-------	-------	-------	-------	-------	--	--	--	--	--

1	2	3	4	5	6	7	8	9	10	11
A	B	C	D	B	C	A	B	C	A	D

# Lempel–Ziv (1977)

- Osumien etsiminen koko tekstin seasta työlästä
- Yleensä ylläpidetään liukuvaa ikkunaa: esim. viimeiset 32 KB dataa

- Pakkausalgoritmeja sivuavia
- Ad hoc -henkisiä