

Semestrální projekt KOD RLE, BWT a LZW

Petr Kašpar, KAS265

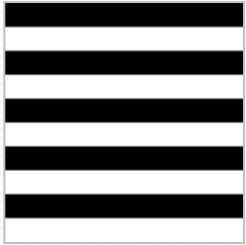
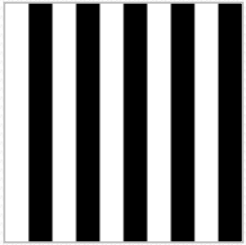
Obsah

- Run-length encoding (RLE)
- Burrows-Wheeler transformace (BWT)
 - Ukázka BWT/RLE
- Lempel-Ziv-Welch (LZW)
 - Ukázka komprese
 - Ukázka dekomprese
- Výsledky experimentu
- Zhodnocení

Run-length encoding (RLE)

- Bezeztrátová komprese
- Posloupnosti stejných symbolů
 - Kódováno do tvaru (znak, délka posloupnosti)
- Účinnost komprese
- Použití
 - PCX, JPEG
- Ukázka
 - **aaabbbbbbaabbbbaa**
 - **a3b6a2b5a2**

RLE vs. kompozice obrazu

Vstupní obraz (PCX)	Velikost
	1,9kB
	6,7kB

Burrows-Wheeler transformace

- BWT
- Pomocný algoritmus
- Záměna pořadí symbolů
- Neprovádí kompresi

Vstupní slovo

^BANANA

Burrows-Wheeler transformace

- BWT
- Pomocný algoritmus
- Záměna pořadí symbolů
- Neprovádí kompresi

Vstupní slovo

^BANANA♠

Burrows-Wheeler transformace

- BWT
- Pomocný algoritmus
- Záměna pořadí symbolů
- Neprovádí kompresi

Vstupní slovo	Všechny rotace
\wedge BANANA♠	\wedge BANANA♠ ♠ \wedge BANANA A♠ \wedge BANAN NA♠ \wedge BANA ANA♠ \wedge BAN NANA♠ \wedge BA ANANA♠ \wedge B BANANA♠ \wedge

Burrows-Wheeler transformace

- BWT
- Pomocný algoritmus
- Záměna pořadí symbolů
- Neprovádí kompresi

Vstupní slovo	Všechny rotace	Seřazené rotace
\wedge BANANA♠	\wedge BANANA♠ ♠ \wedge BANANA A♠ \wedge BANAN NA♠ \wedge BANA ANA♠ \wedge BAN NANA♠ \wedge BA ANANA♠ \wedge B BANANA♠ \wedge	ANANA♠ \wedge B ANA♠ \wedge BAN A♠ \wedge BANAN BANANA♠ \wedge NANA♠ \wedge BA NA♠ \wedge BANA \wedge BANANA♠ ♠ \wedge BANANA

Burrows-Wheeler transformace

- BWT
- Pomocný algoritmus
- Záměna pořadí symbolů
- Neprovádí kompresi

Vstupní slovo	Všechny rotace	Seřazené rotace	Výstup
\wedge BANANA♠	\wedge BANANA♠ ♠ \wedge BANANA A♠ \wedge BANAN NA♠ \wedge BANA ANA♠ \wedge BAN NANA♠ \wedge BA ANANA♠ \wedge B BANANA♠ \wedge	ANANA♠ \wedge <u>B</u> ANA♠ \wedge BAN <u>N</u> A♠ \wedge BANAN <u>N</u> BANANA♠ \wedge <u>_</u> NANA♠ \wedge B <u>A</u> NA♠ \wedge BAN <u>A</u> \wedge BANANA♠ <u>♠</u> ♠ \wedge BANANA <u>A</u>	BNN \wedge AA♠A

Ukázka RLE + BWT

- Vstupní slovo
 - statistika statistika statistika statistika statistika statistika
- BTW
 - aaaakkkkkttttttkttttttttttiiii♠ iiiiiisssssssssssssssaaaaaaa
- RLE
 - a5k5t6k1t12i6♠1 5i6s12a7

Lempel-Ziv-Welch (LZW)

- Slovníková bezeztrátová metoda
- Počáteční inicializace slovníku
- Přidávání víceznakových řetězců
- Maximální velikost slovníku 4095
- Použití
 - GIF
 - PDF

LZW – ukázka komprese

Vstup: B A B A A B A A A

Výstup:

LZW – ukázka komprese

Vstup: B A B A A B A A A

- Inicializace slovníku

Adresa	Slovo
1	A
2	B

Výstup:

LZW – ukázka komprese

Vstup: **B A** B A A B A A A

- B je ve slovníku
 - BA není ve slovníku, vložíme BA na adresu 3, na výstup adresa prefixu (B) → 2

Adresa	Slovo
1	A
2	B
3	BA

Výstup: 2

LZW – ukázka komprese

Vstup: B A B A A B A A A

- A je ve slovníku
 - AB není ve slovníku, vložíme AB na adresu 4, na výstup adresa prefixu (A) → 1

Adresa	Slovo
1	A
2	B
3	BA
4	AB

Výstup: 2 1

LZW – ukázka komprese

Vstup: B A B A A B A A A

- BA je ve slovníku
 - BAA není ve slovníku, vložíme BAA na adresu 5, na výstup adresa prefixu (BA) → 3

Adresa	Slovo
1	A
2	B
3	BA
4	AB
5	BAA

Výstup: 2 1 3

LZW – ukázka komprese

Vstup: B A B A A B A A A

- AB je ve slovníku
 - ABA není ve slovníku, vložíme ABA na adresu 6, na výstup adresa prefixu (AB) → 4

Adresa	Slovo
1	A
2	B
3	BA
4	AB
5	BAA
6	ABA

Výstup: 2 1 3 4

LZW – ukázka komprese

Vstup: B A B A A B A A A

- A je ve slovníku
 - AA není ve slovníku, vložíme AA na adresu 7, na výstup adresa prefixu (A) → 1

Adresa	Slovo
1	A
2	B
3	BA
4	AB
5	BAA
6	ABA
7	AA

Výstup: 2 1 3 4 1

LZW – ukázka komprese

Vstup: B A B A A B A A A

- AA je ve slovníku
 - žádný další znak již není, na výstup adresa (AA) → 7

Adresa	Slovo
1	A
2	B
3	BA
4	AB
5	BAA
6	ABA
7	AA

Výstup: 2 1 3 4 1 7

LZW – ukázka dekomprese

Vstup: 2 1 3 4 1 7

- Inicializace slovníku

Adresa	Slovo
1	A
2	B

Výstup:

LZW – ukázka dekomprese

Vstup: ② 1 3 4 1 7

- Adresa 2 je ve slovníku
 - na výstup slovo na adrese
2 → B

Adresa	Slovo
1	A
2	B

Výstup: B

LZW – ukázka dekomprese

Vstup: 2 1 3 4 1 7

- Adresa 1 je ve slovníku
 - na výstup slovo na adrese 1 → A
 - do slovníku vložíme BA na adresu 3

Adresa	Slovo
1	A
2	B
3	BA

Výstup: B A

LZW – ukázka dekomprese

Vstup: 2 1 3 4 1 7

- Adresa 3 je ve slovníku
 - na výstup slovo na adrese 3 → BA
 - do slovníku vložíme AB na adresu 4

Adresa	Slovo
1	A
2	B
3	BA
4	AB

Výstup: B A B A

LZW – ukázka dekomprese

Vstup: 2 1 3 **4** 1 7

- Adresa 4 je ve slovníku
 - na výstup slovo na adrese 4 → AB
 - do slovníku vložíme BAA na adresu 5

Adresa	Slovo
1	A
2	B
3	BA
4	AB
5	BAA

Výstup: B A B A A B

LZW – ukázka dekomprese

Vstup: 2 1 3 4 **1** 7

- Adresa 1 je ve slovníku
 - na výstup slovo na adrese 1 → A
 - do slovníku vložíme ABA na adresu 6

Adresa	Slovo
1	A
2	B
3	BA
4	AB
5	BAA
6	ABA

Výstup: B A B A A B A

LZW – ukázka dekomprese

Vstup: 2 1 3 4 1 **7**

- Adresa 7 není ve slovníku
 - do slovníku na adresu 7 předchozí výstup + jeho první znak → AA
 - na výstup slovo na adrese 7 → AA

Adresa	Slovo
1	A
2	B
3	BA
4	AB
5	BAA
6	ABA
7	AA

Výstup: B A B A A B A A A

Experiment – SMS zprávy

Text #	Originál	RLE+BWT	LZW	WinRAR	LZW kompr. poměr
1	150	254	112	184	1,34 : 1
2	20	38	18	93	1,11 : 1
3	90	152	73	151	1,23 : 1
4	19	38	18	92	1,06 : 1
5	50	90	43	123	1,16 : 1
6	92	158	76	154	1,21 : 1
7	141	142	109	181	1,29 : 1
8	63	64	55	136	1,15 : 1
9	45	86	41	118	1,10 : 1
10	34	64	31	107	1,10 : 1
11	714	1022	618	497	1,16 : 1

Použitá literatura

- Lempel-Ziv-Welch [online]. 2009 [cit. 2009-04-19]. Dostupný z WWW: <<http://en.wikipedia.org/wiki/Lempel-Ziv-Welch>>.
- Run-length encoding [online]. 2009 [cit. 2009-04-19]. Dostupný z WWW: <http://en.wikipedia.org/wiki/Run-length_encoding>.
- Burrows-Wheeler transform [online]. 2009 [cit. 2009-04-19]. Dostupný z WWW: <http://en.wikipedia.org/wiki/Burrows-Wheeler_transform>.
- KOJECKÝ, Marek. Komprese binárních textových a grafických dat. [s.l.], 2005. 46 s. Bakalářská práce.
- Lempel-Ziv-Welch (LZW) Compression Algorithm [online]. [2002] [cit. 2009-04-19]. Dostupný z WWW: <faculty.kfupm.edu.sa/ICS/jauhar/ics202/Unit32_LZW.ppt>.

Děkuji za pozornost