

Huffmanovo kódování

je příkladem bezztrátového kompresního algoritmu. Nejčastnějším znakům přiřazuje kratší kódovací sekvenci, čímž šetří místo. Je detailně popsán např. na [anglické wiki](#), existuje [online applet](#), který provede rozbor na strom libovolné zadané věty.

Příklad komprese textu

Mějme větu: **TENTO TEXT JE JEN TEST**

V klasickém kódování ASCII zabere 22 B neboli 176 b.

V Huffmanově kódování pouze 60 b tedy necelých 8 B.

Postup

1. **Zjistíme četnost** jednotlivých znaků:

Znak	T	E	SPC	N	J	O	X	S
Četnost	6	5	4	2	2	1	1	1

2. **Vytvoříme strom** četností: Vezmeme dva znaky s nejnižší četností, vytvoříme z nich nový „složený znak (dvoznak)“, jehož četnost odpovídá součtu četností znaků ve dvojici. To opakujeme tak dlouho, dokud nezbude jen jediný „složený znak“ ze všech znaků obsažených ve zprávě.
3. Zpětně **procházíme vzniklý strom** a na každé úrovni **přidělujeme vždy symboly 1 a 0** tak, že větev/znak s vyšší četností dostane 1 a větev/znak s nižší četností 0.

Na příkladu

1. Nejnižší četnost mají znaky X,S, proto vznikne dvoznak (X,S) s četností 2:

Znak	T	E	SPC	N	J	(X,S)	O
Četnost	6	5	4	2	2	2	1

2. Další dvojice v pořadí je O a dvoznak (X,S):

Znak	T	E	SPC	((X,S),O)	N	J
Četnost	6	5	4	3	2	2

3. Nyní mají nejnižší četnost znaky N a J, vzniká z nich dvoznak:

Znak	T	E	SPC	(N,J)	((X,S),O)
Četnost	6	5	4	4	3

4. V dalším kroku budeme slučovat trojznak XSO a dvoznak NJ, který bude mít hodnotu 7:

Znak	((N,J), ((X,S),O))	T	E	SPC
Četnost	7	6	5	4

5. Dále slučujeme mezeru a E:

Znak	(E,SPC)	((N,J), ((X,S),O))	T
Četnost	9	7	6

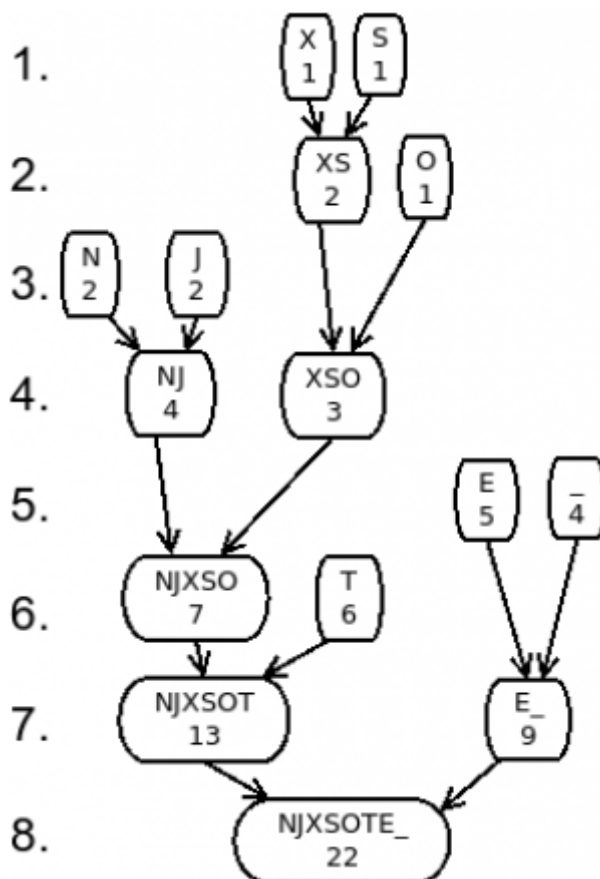
6. V předposledním kroku bude složený znak NJXSO spojen s T:

Znak	(((N,J),(X,S),O)),T)	(E,SPC)
Četnost	13	9

7. Při posledním složení už vznikne jediný složený znak, tedy kompletní strom: - V předposledním kroku bude složený znak NJXSO spojen s T:

Znak	((((N,J),(X,S),O)),T),(E,SPC))
Četnost	22

Výše uvedený řetězec lze zapsat také formou stromu



Při zpětném sestavení přidělujeme 1 a 0

krok:	1	2	3	4	5	Celkem
T	1	0				10
E	0	1				01
SPC	0	0				00
N	1	1	1	1		1111
J	1	1	1	0		1110
O	1	1	0	0		1100
X	1	1	0	1	0	11010
S	1	1	0	1	1	11011

[online zpracování stromu](#)

Zakódovaný text

100111111011000010011101010001110010011100111110010011101110

Dekódování probíhá **sekvenčně** náhradou za znak z tabulky, vždy jakmile to nalezená sekvence umožňuje:

10	01	1111	10	1100	00	10	01	11010	10	00	1110	01	00	1110	01	1111	00	10	01	11011	10
T	E	N	T	O		T	E	X	T		J	E		J	E	N		T	E	S	T

From:

<https://gml.cz/wiki/> - GMLWiki

Permanent link:

<https://gml.cz/wiki/doku.php/informatika:maturita:huffmann>



Last update: **26. 09. 2020, 23.18**