

Aktivnost 3

Stiskanje besedil

Zdaj znamo stiskati slike. Stiskati pa je potrebno tudi druge reči. V tej aktivnosti bomo spoznali enega naprednejših postopkov; ogledali si bomo, kako bi se z njim lotili stiskanja besedila.

Namen

Otroci spoznajo osnovo idejo postopkov za stiskanje besedil Lempel-Ziv-Welch. Splošnejši namen je spoznavanje različnih načinov kodiranja podatkov.

Trajanje

Ena ura

Potrebščine

Učni listi so kar v tem besedilu, strani od 3 do 6.

Stiskanje besedil

Deli besedila – zaporedja črk, besede ali celo več besed – se pogosto ponavljajo. Zapis besedila lahko skrajšamo tako, da vsakič, ko bi morali napisati, kaj, kar smo že napisali, le pokažemo na prejšnji zapis.

Kaj pa piše tule? (Nariši na tablo.)

AN B PET PODG

Tole je naslov neke dobre knjige. Katere?

V PUŠČAVI IN GO

Je v tvojem razredu morda kakšna

BAR A?

Zdaj pa se nauči še sam zastavljati takšne uganke. Kako bi zapisal naslov knjige PEDENJPED? Kaj pa VIKING VIKE? Kaj pa beseda VODOVOD? Izmisli si še sam kako besedo ali stavek, ki ga je mogoče dobro stisniti.

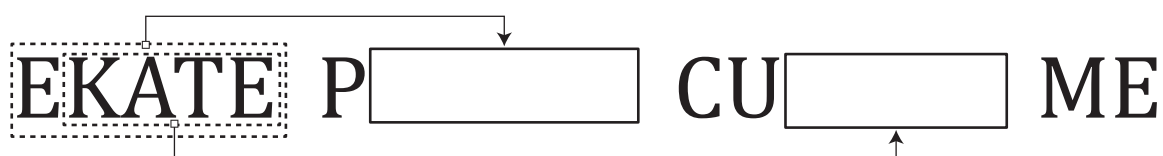
Učencem razdeli učne pole (preostanek tega besedila, razen zadnje strani).

Prepletanje

Kako bi stisnili tole vrstico stare dobre izštevankе?

EKATE PEKATE CUKATE ME

To ni več tako preprosto: ko napišemo PEKATE, uporabimo cel EKATE, pri CUKATE, pa le KATE, takole:



Naprej je na srečo veliko lažje. Bi znal nadaljevati? Obkroži enake dele, prečrtaj ponovljene dele in dodaj puščice!

EKATE PEKATE CUKATE ME

FIBE FABE DOMINE

EKTUM PEKTUM KUFER ŠTUC

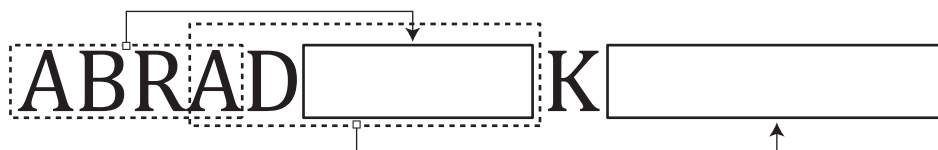
KVINTE KVANTE FINGER PUC

Poskusi tudi

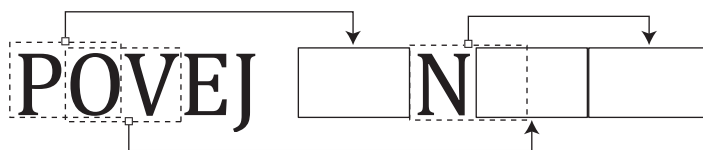
ŽIGA ŽAGA POJE ŽAGA, ROM POM POM KLADIVO.

Branje stisnjene besedila

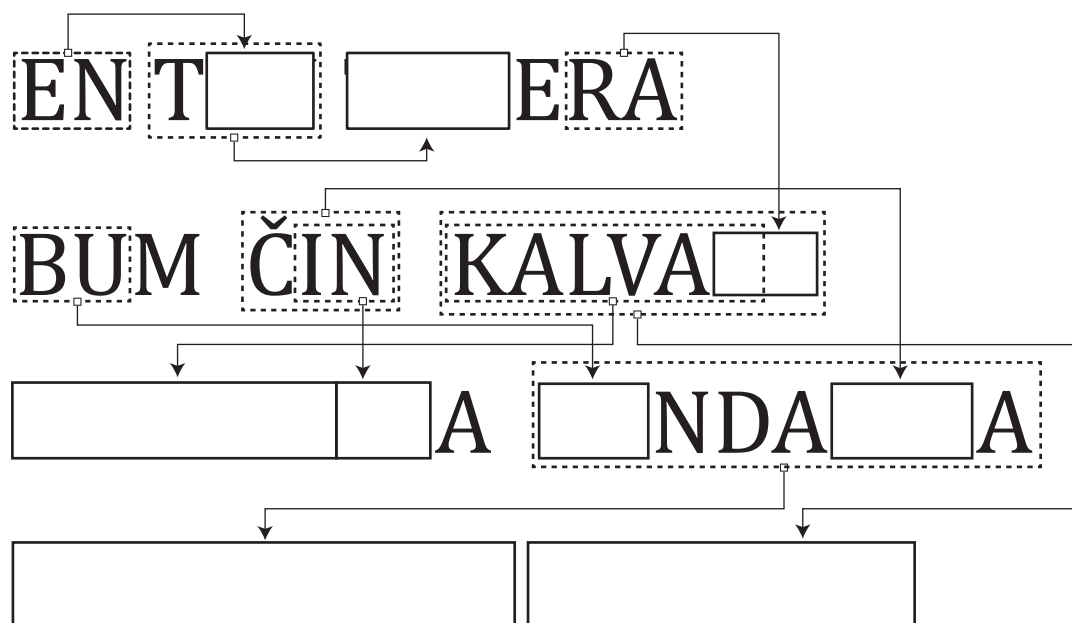
Tudi branje lahko postane kar zapleteno. Znaš razvložiti tole čarovnijo?



Odlično ti gre! Še ena zavozlanka:



In še zadnja, za najbolj neustrašne:



Kako to zapisati?

Kako bi to predstavili v računalniku? Vsakič, ko se želimo sklicati na prejšnje besedilo, v oklepaju povejmo, koliko znakov je potrebno prepisati in odkod. Za primer vzemimo

V PUŠČAVI IN GO

in ga prerišimo tako, da bomo lahko šteli črke:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
V	P	U	Š	Č	A	V	I	I	N	G	O								

Ko pridemo do šestnajstega znaka, vstavimo pet znakov od petega mesta naprej. Torej bomo zapisali: V PUŠČAVI IN GO[5, 5].

AN B PET PODG

bomo zapisali kot AN B[2, 1] PET PODG[2, 1]: za B-jem sta dva znaka, prepisovanje začnemo kar pri prvem znaku besedila; dva znaka, začnši od prvega, sta AN. Enako storimo za G-jem. Lahko pa bi (razmisli!) rekli tudi AN B[2, 4] PET PODG[2, 5].

EKATE P CU ME

se prepíše v EKATE P[5, 1] CU[4, 2]KATE ME: P-ju sledi pet znakov, ki jih prepíšemo z začetka – to je EKATE – in CU-ju štirje znaki, ki jih prepisujemo od drugega naprej, torej KATE.

Za kaj gre?

Čeprav imajo računalniki vedno večje diske, prostora na njih še vedno zmanjkuje, saj želimo nanje shraniti vedno več reči. V računalnikih imamo cele knjižnice, velike zbirke glasbe in filmov, tisoče in tisoče slik... V zadnjem času se vse to bogastvo seli celo na mobilne telefone, kjer smo z diskom veliko bolj na tesnem.

Da bi šlo v isto količino pomnilnika ali diska več podatkov, podatke stiskamo. V prejšnji aktivnosti smo spoznali preprosto postopek stiskanja slik, ta aktivnost kaže osnovno ozadje pogoste metode stiskanja, ki se imenuje stiskanje LZ, po izraelskih profesorjih Zivu in Lempelu, ki sta se ga domislila v sedemdesetih. Večina postopkov stiskanja – vključno s postopkom LZ – temelji na dejstvu, da se določeni vzorci v podatkih ponavljajo, zato se pri ponovitvah le skličemo na že opisani vzorec. Takšno stiskanje se ne uporablja le za besedila, temveč postopek LZW (Lempelu in Zivu se tu pridruži še Welch) uporablja tudi priljubljeni format slik GIF, včasih pa so z njim stisnjene tudi datoteke TIFF in PDF. Uporablja se tudi v formatu ZIP in podobnih.

Nekatere druge metode stiskanja delujejo tako, da znake, ki se pojavljajo pogosteje (na primer črki E in A), napišejo z manj biti, redkejša črka pa z več. Na tem delno temelji Morsejeva abeceda.

S takšnimi postopki lahko besedilo običajno stisnemo na polovico prvotne velikosti.

Naštete metode stiskanja imenujemo brezizgubne, saj je mogoče podatke prebrati ("odstisniti") tako, da spet dobimo prvotno besedilo. Slikovne in zvokovne datoteke je težko stiskati na ta način, saj se vzorci le redko natančno ponovijo. Zanje zato uporabljamo izgubne postopke, ki sliko ali zvok nekoliko poenostavijo. Tako stiskamo na račun kvalitete; pri zvoku, recimo, stiskanje poreže nižje in višje frekvence, tišje tone, zmanjša število različnih frekvenc... Tako lahko, na primer, pri gledanju videov na YouTubeu določimo, kako kvaliteten posnetek želimo; če je naša internetna povezava manj zmogljiva, bomo izbrali bolj stisnjen video, ki pa bo zaradi tega opazno slabši. Prav tako je zvok v obliki MP3 bolj ali manj stisnjen; čim manjša je datoteka, tem slabši je zvok.