

# Aktivnost 1B

---

## Krajši zapis besedil: Huffmanovo kodiranje

### Povzetek

Da zapišemo črko s številko, potrebujemo pet znakov, ne? Ne. Lahko jih potrebujemo tudi manj in več. (Manj *in* več?!)

### Namen

Otroci spoznajo idejo "kodiranja": podatke lahko zapisujemo različno učinkovito, z različno dolžino zapisa. Mimogrede spoznavajo tudi drevesa.

### Predvideni čas izvajanja aktivnosti

Ena šolska ura

### Potrebščine

- Trakci z besedili, ki jih bodo brali otroci. Pripravi nekajkrat več trakcev, kot je otrok, saj bodo hoteli ponavljati igro.
- Vsak otrok potrebuje tudi pisalo
- Otrokom utegne priti prav tudi tabela s črkami, ki je na poli Ujeti Štefan iz prejšnje aktivnosti
- Kreda za risanje dreves (drevesa so velika in zahtevajo kar nekaj kred)

Drevesa za kodiranje, narisana na šolskem dvorišču ali označena na travniku ali drugem primernem mestu. Namesto tega lahko natisnemo tudi liste z drevesi za vsakega otroka.

### Dodatna navodila

Aktivnost je najboljšje izvajati na prostem.

Na tla boš morala narisati drevo; primerna so asfaltna tla pred šolo, risanja pa bo zahtevalo veliko krede. Druga možnost je, da drevesa pripraviš na travniku tako, da v tla zapičiš palice in med njimi povlečeš vrvice, na koncu pa so zastavice z listki. To je primernejše za večje dogodke, saj zahteva več priprav.

Drevesa so na koncu tega opisa aktivnosti in na ločenih listih. Risanje Huffmanovega drevesa je zahtevnejše, zato je na listu narisano na kvadratni mreži. Mreže ne prerisuješ, temveč služi le kot pomoč pri risanju.

Narišeš lahko eno, dva ali tri drevesa.

Huffmanovo drevo, v katerem so različne črke na različni globini, je bistveni del aktivnosti.

V drugem drevesu so vse črke na isti višini. Drevo ustreza kodiranju, kakršnega je uporabljal ujeti Štefan. Za otroke je koristno, da vidijo povezavo med dvojiškim številskim sistemom in drevesi. Iz dreves vidijo tudi, kakšna je razlika med kodiranjem z enako in različno dolžino kod. Iz primerjave dreves jim je "fizično" očitno, da so pri Huffmanovem drevesu pogosti znaki bližje.

Drevo, ki ustreza Morsejevi kodi je zanimivo zaradi Morsejeve kode kot takšne, v sklopu aktivnosti pa je zanimivo, ker kaže primer kodiranja, v katerem je koda določenega znaka lahko prefiks kode drugega znaka. Tako je, na primer, črka I zapisana kot "..", U pa kot "..-". Zaradi tega v Morsejevem kodiranju potrebujemo tudi "presledke" med znaki. Morsejeva koda torej nima le dveh znakov, pike in črte, temveč tudi presledek.

Presodi, ali označiti veje s črkama L in D ali ne. Drevesa bodo preglednejša brez njih. Pri mlajših otrokih jih lahko izpustiš tudi, da bodo vadili levo in desno, vendar to raje stori le pri drevesih s konstantno dolžino zapisa, kjer bo zaradi ene napačne smeri napačna le ena črka. Pri Huffmanovih drevesih bo že ena napaka pokvarila večji del sporočila.

Za vsakega otroka pripravi trakove s sporočili. Otroci naj imajo različna sporočila, sicer bodo vsi hodili po isti poti. Na traku naj bo tudi dovolj prostora, da bodo otroci vanj pisali prebrano sporočilo.

Uporabiš lahko sporočila, objavljena na strani, ali pa sestaviš svoja s pomočjo temu namenjene spletne strani (glej povezavo na spletni strani z opisom aktivnosti). Ker bo tega več listov, si boš delo precej olajšala, če bodo vsi zapisi dolgi enako število vrstic in boš lahko rezala več listov hkrati. (Sporočila, ki so objavljena na strani, so že pripravljena tako.)

Pri Huffmanovem kodiranju lahko uporabiš različico s presledki med znaki ali takšno, v kateri presledkov ni. Druga je veliko poučnejša, vendar ji je nekoliko težje slediti, zato pri mlajših otrocih (npr. prvi razredi OŠ) raje uporabi (tudi?) prvo.

# Kodiranje besedila

## Uvod

Kako si računalniki pošiljajo besedila? Že v prvi aktivnosti smo ob Štefanu, ki je bil ujet v trgovini, videli, da lahko pošiljamo sporočila tako, da vsako črko spremenimo v številko, številko zapišemo po dvojiško in se dogovorimo, kako bomo predstavili ničlo in kako enico. Štefan je prižigal in ugašal luč; namesto tega bi lahko dvigal ali spušča zastavico, mahal z levo ali desno roko.

## Pogovor

Otroke spomni, kaj so počeli v prejšnji aktivnosti, zapisovanje besedil. Koliko znakov (dvigov zastavice, listkov obešenih na vrv, tleskov...) so potrebovali za zapis vsake črke?

Vprašaj otroke, ali bi šlo z manj kot petimi znaki. Ker vedo, da je črk 25 in ker so se v prvi aktivnosti naučili, koliko bitov je potrebnih za prenos tako velikih števil, bodo vedeli, da potrebujejo pet znakov.

Spomni jih, da tudi računalniki shranjujejo in pošiljajo besedilo na podoben način, le da si namesto iztegovanja rok pošiljajo električne signale. Če bi uspeli namesto petih znakov zapisati vsako črko z, recimo, štirimi, bi komunikacija (recimo internet) delovala hitreje. Vprašaj jih, ali so prepričani, da ne bi šlo.

Povej jim, da imajo (skoraj) prav: ne moremo si izmisliti zapisovanja, v katerem bi vsako črko opisali s štirimi znaki. Lahko pa naredimo tole: nekatere črke zapišemo z manj kot petimi, vendar bomo morali zaradi tega nekatere opisati z več kot petimi.

## Branje besedil

Otrokom razdeli listke s stavki. Če uporabljaš tako Huffmanova drevesa kot drevesa s konstantno dolžino zapisa, pazi, da bodo otroci uporabljali prave listke za prava drevesa: listke s konstantno dolžino prepoznaš po tem, da so znaki urejeni v skupine po pet. Pomagaš jim lahko tudi tako, da sporočila za različna drevesa tiskaš v različnih barvah ali na papir različnih barv ter označiš drevesa z isto barvo.

Razloži jim, kako brati besedila: postavimo se na začetek drevesa. Beremo znake na listu (L in D za Huffmanova drevesa in drevesa s fiksno dolžino kode; pike in črte za Morsejevo drevo) in skladno z njimi izbiramo leve in desne veje v drevesu, dokler ne pridemo do konca. Ko pridejo do lista drevesa, si zapišejo črko v njem. Nato gredo spet na začetek drevesa in berejo naslednji znak.

Posebej pri Huffmanovem drevesu bo branje preprostejše, če se otroci razdelijo v pare: eden bere ničle in enice z lista, drugi pa hodi po drevesu in sporoča znake, do katerih je prišel.

Če aktivnost izvajaš v razredu (škoda!) lahko drevesa natisneš na papir in otroci rešujejo nalogo na papirju.

Pripravljena besedila vsebujejo naslednje stavke, ki jih je prijazno prispevala Tjaša Medved:

- |    |                            |    |                            |
|----|----------------------------|----|----------------------------|
| 1  | Bil sem v živalskem vrtu.  | 18 | Na vrtu dela moja mama.    |
| 2  | Zunaj sije rumeno sonce.   | 19 | Babica skuha dobro kosilo. |
| 3  | Moj brat igra vaterpolo.   | 20 | Najraje jem sadno kupo.    |
| 4  | Letos bom star devet let.  | 21 | Brat Jure je zaljubljen.   |
| 5  | Jutri bom stara tri leta.  | 22 | Kmalu bom prišel domov.    |
| 6  | Rad hodim v osnovno šolo.  | 23 | Všeč mi je rumena barva.   |
| 7  | V svoji sobi imam knjige.  | 24 | Čokoladna torta je dobra.  |
| 8  | Na poletni šoli je super.  | 25 | Avto je zelo lep in velik. |
| 9  | Najrajši imam sladoled.    | 26 | Rada jem čokolado Milka.   |
| 10 | Računalništvo je zakon.    | 27 | Moja mami pije kokakolo.   |
| 11 | Višnja gora je lep kraj.   | 28 | Anja poleti rada plava.    |
| 12 | Jure in Ana gresta v park. | 29 | Videla sem morskega psa.   |
| 13 | Kaja in Jan gresta domov.  | 30 | Sončni zahod je čaroben.   |
| 14 | Na praznovanju je super.   | 31 | Rada se igram s sestrico.  |
| 15 | V naši hiši je zelo mrzlo. | 32 | Rad skačem v morsko vodo.  |
| 16 | Na mizi je šopek vrtnic.   | 33 | Sosed Jurček je prijazen.  |
| 17 | Počitnice so prekratke.    |    |                            |

### Pisanje besedil

Otroci lahko podobna besedila pišejo tudi sami. Prebrati jih lahko dajo sošolcem, v posebno veselje pa jim bo, če jih bo bral učitelj.

## Pogovor

Huffmanovo kodiranje črk ne zapisuje z enakim številom znakov (bitov), torej pet. Črke, ki so pogostejše, opiše z manj znaki, redkejše z več. Če se ozremo po drevesu, vidimo, da gre takole:

Število bitov (globina v drevesu)	Znaki
3	E, presledek
4	T, R, N, O, I, A
5	Z, V, D, P, L, J, M, K, S
6	G, C, B, U
7	H, F, Č
8	Ž, Š

Čeprav nekateri znaki zahtevajo celo osem bitov, sta le dva in se pojavljata redko. Po drugi strani pa veliko pridobimo s tem, ko pogoste znake zapisujemo z malo biti.

V Huffmanovem in drugih podobnih kodiranjih so kode sestavljene tako, da koda nobenega znaka ne predstavlja začetka kode kakega drugega znaka. Znak N zapišemo kot 1000, zato se noben drug znak ne začne z 1000. Da je tako, se hitro prepričamo iz drevesa: če bi imeli, recimo, znak 10001, bi se ta nahajal pod črko N. Tam pa ga seveda ni, saj so črke vedno v listih drevesa.

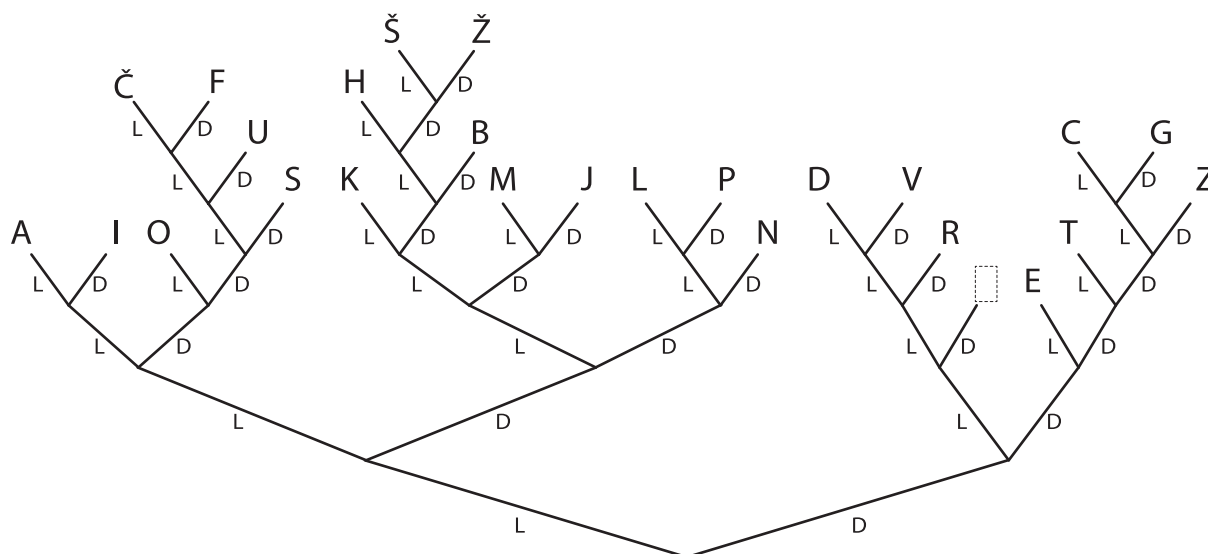
Za ilustracijo so tu kode vseh znakov v drevesu:

Z	00000	M	10101
G	000010	B	10100
C	000011	Ž	10110100
T	00001	Š	10110101
E	001	H	1011011
presledek	010	K	10111
R	0110	S	11000
V	01110	U	110010
D	01111	F	1100110
N	1000	Č	1100111
P	10010	O	1101
L	10011	I	1110
J	101100	A	1111

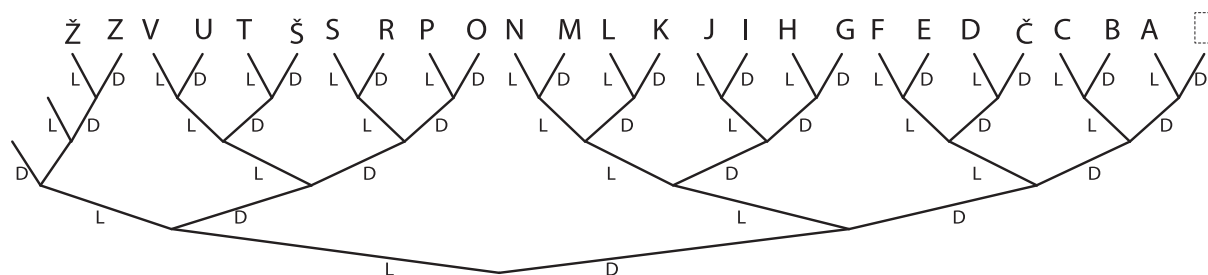
Očitno se noben znak ne začne s kodo kakega drugega znaka.

Morsejevo kodiranje nima te lastnosti, zato je potrebno med znaki delati presledke, s katerimi povemo, da je znaka konec in moramo spet na začetek drevesa.

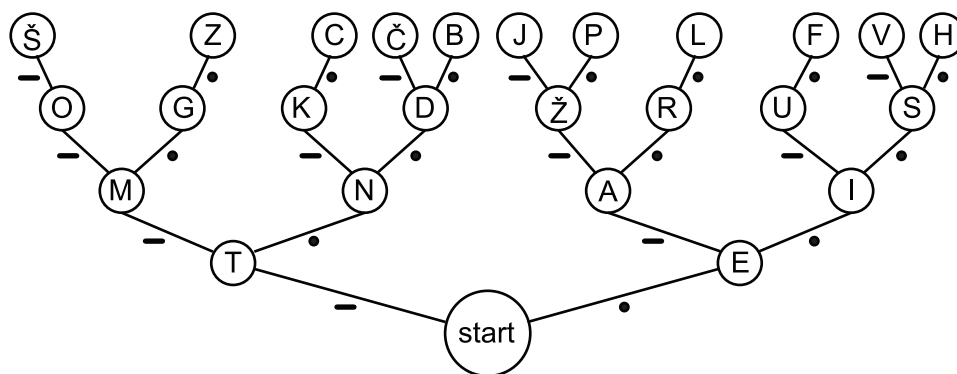
## Huffmanovo drevo za slovenščino



## Kodiranje s konstantno dolžino (zaporedne številke)



## Morsejeva koda



## Za učitelje: za kaj gre?

Huffmanovo kodiranje je eden od načinov stiskanja podatkov. Dva druga spoznavamo v aktivnosti stiskanje slik in stiskanje besedil.

Huffmanovo kodiranje temelji na teoriji informacij. Če imamo dva dogodka, ki sta enako verjetna, za opis izida poskusa potrebujemo en bit. Primer je met kovanca, pri čemer bomo z 0 opisali, na primer, grb in z dva številko. Podobno za opis števila med 0 in 31 potrebujemo pet bitov – vendar le, če se vsa števila enako verjetna.

Zdaj vzemimo par različno verjetnih dogodkov. Opazujemo, na primer, vlak, ki včasih iztiri, a zelo redko – recimo enkrat v tisoč vožnjah. Če bi podobno kot pri metu kovanca opisovali možna izida vožnje z 0 in 1 (0 pomeni, da vlak ne iztiri in 1, da iztiri), bi za opis tisoč voženj potrebovali tisoč bitov. Ker pa vemo, da so iztiritve redke, bomo "evidenco iztiritov" raje vodili tako, da bomo zapisali zaporedne številke voženj, v katerih je vlak iztiril. Ali, še boljše, število voženj med dvema zaporednima iztiritvama. Število voženj bomo lahko zapisali s kakimi desetimi biti (oziroma, v praksi, malo več, zaradi tehničnih detajlov), saj deset bitov zadošča za zapis števil med 0 in 1023. Na vsakih 1000 voženj bomo torej potrebovali 10 bitov. Posamično vožnjo torej opišemo z 0,01 bita.

Huffmanovo kodiranje se trudi vsako besedilo opisati s približno toliko biti, kolikor jih dejansko zahteva. Vsaka črka predstavlja "dogodek", dolžina njegovega zapisa pa pogostosti tega dogodka.