Zapis podatkov v računalniku - 3. del

Zapis znakov

Pisni znaki so v računalniku kodirani z

Dvojiško kodiraj naslednje znake:

Znak: A B C D

Dvojiška koda:

Ali lahko te znake kodiramo še kako drugače?

Možnosti za kodiranje pisnih znakov je

Poenotenje kodiranja pisnih znakov je pomembno zaradi

Zato so uvedli

Standard za kodiranje pisnih znakov se imenuje  (American Standard Code for Information Interchange, kar pomeni ameriški standardni zapis za izmenjavo informacij).

Po standardu ASCII kodiramo znake z **.** Na ta način lahko predstavimo različnih znakov. Kod ASCII predpisuje le znakov. Neposredno se ta kod lahko uporablja le v **.** Pri nas so problem **.**

Pisave drugih dežel povečini vsebujejo **,** ki jih ASCII ne pozna.

Žal pa tudi ASCII z dopolnitvami ne zajema vseh različnih pisav. Zato se v računalništvu uveljavlja nov standard imenovan **.** V njem se uporablja za zapis niz s biti. Z njim je možno zapisati znakov to je 2.

Unicode

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1a |

ASCII

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |   |   |   |   |   |   |   | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1a |

**Naloga 1:** Zapiši svoje ime v ASCII standardu.

**Naloga 2:** S pomočjo ASCII tabele zakodiraj besedo ANANAS!

****

**Naloga 3:** Velike črke slovenske abecede predstavimo z 0 in 1.

a) Koliko bitov informacije potrebujemo za zapis teh znakov? Račun:

b) Izdelajte kodno tabelo za vseh 25 črk tako, da bo zapis najkrajši.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Koda  | Koda  | Koda  | Koda  | Koda  |
| A  | E  | J  | O  | T  |
| B  | F  | K  | P  | U  |
| C  | G  | L  | R  | V  |
| Č  | H  | M  | S  | Z  |
| D  | I  | N  | Š  | Ž  |

**Naloga 4:**

Kateri standard za zapis znakov (kodna preglednica) določa največ znakov?

**a)** **[ ]  ASCII**

**b) [ ]  UNICODE**

**c) [ ]  ISO 8859-2**

**d) [ ]  LATIN 14**

**e) [ ]  CODE PAGE 1250**

**Naloga 5:** V 8-bitni kodni tabeli ASCII ima znak A kodo 65. V naslednjem nizu enic in ničel so zapisani znaki, kodirani po tabeli ASCII. Kaj je zapisano v nizu enic in ničel?

1000001010000100100001001000001

1. **[ ]  ABAK**
2. **[ ]  ABBA**
3. **[ ]  BABA**
4. **[ ]  MATURA**
5. **[ ]  2014**

**Naloga 6:** Najdaljša slovenska beseda (po SSKJ) je: DIALEKTIČNOMATERIALISTIČEN.

1. **Koliko zlogov zavzame beseda v pomnilniku, če je kodirana v skladu z razširjeno kodno tabelo ASCII?**

1. **Peter želi zapisati besedo s kitajskimi pismenkami. Kako imenujemo standard, ki ga bo uporabil za zapis?**

**Naloga 7:** V kodni tabeli ASCII ima znak B kodo 66(10). Napišite kodo znaka K v tej tabeli.

**Naloga 8:** Znak ☺ ima v kodni tabeli UNICODE kodo 263A(16). V HTML-ju lahko ta znak zapišemo z delcem oblike &#n;, pri čemer je n koda znaka, zapisana v desetiškem sestavu. Določite število n.

**Naloga 9:** Marko, Metka in Miha so pri informatiki dobili nalogo, da izdelajo kodno tabelo za prvih deset velikih črk slovenske abecede. Spodaj so prikazane njihove kodne tabele.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Markova kodna tabela |  | Metkina kodna tabela |  | Mihova kodna tabela |
| A | 0000 |  | A | 00 |  | A | 0 |
| B | 0001 |  | B | 01 |  | B | 1 |
| C | 0010 |  | C | 100 |  | C | 10 |
| Č | 0011 |  | Č | 110 |  | Č | 11 |
| D | 0100 |  | D | 1010 |  | D | 100 |
| E | 0101 |  | E | 1110 |  | E | 101 |
| F | 0110 |  | F | 10110 |  | F | 110 |
| G | 0111 |  | G | 11110 |  | G | 111 |
| H | 1000 |  | H | 101110 |  | H | 1000 |
| I | 1001 |  | I | 111110 |  | I | 1001 |

Ali lahko z vsemi tremi kodnimi tabelami enolično zakodiramo besedo FIGA? Enolično kodiranje pomeni, da če besedo kodiramo, jo moremo enolično dekodirati. Utemeljite odgovor.

Katero kodno tabelo bi bilo najbolj smiselno uporabiti za zapis besede ABBA? Utemeljite odgovor.

Z izbrano kodno tabelo iz prejšnjega vprašanja zapišite besedo ABBA.

**Naloga 10:** Butalski župan se je odločil, da bodo v Butalah uporabljali samo velike tiskane črke slovenske abecede. Namesto arabskih bodo uporabljali rimske številke, saj so Butale obstajale že v rimskih časih in še prej. Od ločil bodo uporabljali le piko, vejico, klicaj, vprašaj in presledek. Odločil se je še, da bodo kode vseh črk, številk in ločil enako dolge.

Najmanj koliko bitov bi po tej kodni tabeli potrebovali za nestisnjen zapis besedila: »BUTALE SO NAJLEPŠI KRAJ NA SVETU!«?

**Naloga 11:**

Za domačo nalogo je moral Peter Zmeda sestaviti kodno tabelo za predstavitev sorodnikov

*oče*, *mama*, *sin*, *hči*, *brat*, *sestra*, *babica*, *dedek*, *teta*, *stric*, *bratranec*, *sestrična*.

Uporabil je dve različni kodiranji, in sicer je v prvem primeru za kodiranje uporabil dva različna znaka, v drugem pa tri različne znake. V obeh primerih so bile kode kar se da kratke, a še vedno nedvoumne. Poleg tega so bile vse kode v vsakem od kodiranj po enako dolge.

V pomoč za spodnje odgovore si lahko izpolnite kodni tabeli za vsako od kodiranj:

|  |  |
| --- | --- |
| Kodna tabela z dvema različnima znakoma: | Kodna tabela s tremi različnimi znaki: |
|

|  |  |
| --- | --- |
| Podatek | Koda |
| oče |  |
| mama |  |
| sin |  |
| hči |  |
| brat |  |
| sestra |  |
| babica |  |
| dedek |  |
| teta |  |
| stric |  |
| bratranec |  |
| sestrična |  |

 |

|  |  |
| --- | --- |
| Podatek | Koda |
| oče |  |
| mama |  |
| sin |  |
| hči |  |
| brat |  |
| sestra |  |
| babica |  |
| dedek |  |
| teta |  |
| stric |  |
| bratranec |  |
| sestrična |  |

 |

Kako dolga je posamezna koda v primeru dveh različnih znakov? Utemeljite, zakaj ne more biti krajša.

Kako dolga je posamezna koda v primeru treh različnih znakov? Utemeljite, zakaj ne more biti krajša.

Seveda v splošnem ni potrebno, da so vse kode enako dolge. Na kaj mora Peter pri tem paziti, če se odloči za takšno kodiranje, da bo lahko vedno enolično dekodiral besedilo?

**Naloga 12:**

Pri zgoščevanju niza znakov s Huffmanovim algoritmom je Peter Zmeda dobil te kode:

za črko A kodo 1

za črko B kodo 0

za črko C kodo 01

za črko D kodo 110

za črko E kodo 010

Izkaže se, da niz 10101010 lahko predstavlja več kot eno zaporedje črk, – pravimo, da je Petrova koda dvoumna. Zapišite vsaj dve zaporedji, ki jih lahko predstavlja.

Kako bi odpravili dvoumnost Petrove kode?

**Naloga 13:**

Recimo, da imamo zdaj opravka z 8-bitno kodno tabelo ASCII, v kateri ima znak A desetiško kodo 65. Katero ime je zapisano s tem zaporedjem bitov:

01010100 01101001 01101101

A [ ]  Tim

B [ ]  Taj

C [ ]  Maj

D [ ]  Gaj

Peter Zmeda si je izmislil novo 7-bitno kodno tabelo za slovenske črke, v kateri so črke urejene tako, da je koda prve črke manjša od kode druge črke, če je prva črka po abecedi pred drugo črko. V njej ima velika črka B kodo 12.

Kakšna je v njegovi tabeli koda velike črke N?