

Osnove programiranja v jeziku Python za neračunalničarje

Miha Moškon

19. julij 2020

Osnutek

Kazalo

1	Uvod	5
1.1	Računalništvo	5
1.2	Zakaj se učiti programiranja	6
1.3	Računalništvo in kemija (in druge vede)	6
1.4	Kaj je programiranje?	7
1.5	Zakaj Python?	8
1.6	Kaj me čaka na koncu knjige?	8
2	Spoznavanje z okoljem	11
2.1	Izbira in namestitev okolja	11
2.2	Okolje IDLE	11
2.3	Ukazna vrstica	12
2.4	Podatkovni tipi	13
2.5	Funkcije	14
2.6	Spremenljivke	15
2.7	Pisanje programov	17
2.8	Funkcija <code>print</code>	19
2.9	Funkcija <code>input</code>	20
2.10	Pretvarjanje med podatkovnimi tipi	20
2.11	Pisanje komentarjev	22
3	Pogojni stavek	23
3.1	Zakaj pogojni stavki?	23
3.2	Osnovna oblika stavka <code>if</code>	24
3.3	Kaj je pogoj?	24
3.3.1	Primerjalni operatorji in podatkovni tip <code>bool</code>	24
3.3.2	Operatorja vsebovanosti	27
3.3.3	Združevanje rezultatov primerjanja	28
3.4	Veja <code>else</code>	29
3.5	Veja <code>elif</code> in gnezdenje stavkov <code>if</code>	31
4	Zanka <code>while</code>	35

4.1	Kaj so zanke?	35
4.2	Zanka while	35
4.3	Štetje z zanko while	37
4.4	Stavek +=	37
4.5	Mrtva zanka	38
4.6	Stavek break	38
4.7	Veja else	41
5	Seznami in metode	45
5.1	Sekvenčni podatkovni tipi	45
5.2	Kaj so sezname?	45
5.3	Indeksiranje seznamov	46
5.4	Operatorji nad sezname	47
5.5	Spreminjanje in brisanje elementov seznama	48
5.6	Vgrajene funkcije nad sezname	49
5.7	Metode	49
5.8	Dodajanje elementov	50
5.9	Branje seznamov iz ukazne vrstice	51
5.10	Sortiranje seznamov	52
5.11	Seznami seznamov	54
5.12	Generiranje seznamov s funkcijo range	56
5.13	Režine	59
5.14	Indeksiranje nizov	61
5.15	Sprehajanje čez seznane	63
6	Zanka for	65
6.1	Sprehajanje čez seznane z zanko for	65
6.2	Sprehajanje s funkcijo range in sprehajanje čez indekse	67
6.3	Sprehajanje čez elemente ali čez indekse?	67
6.4	Spreminjanje elementov seznama z zanko for	69
6.5	Zanka for ali zanka while ?	70
6.6	Stavek break	70
6.7	Veja else	71
6.8	Gnezdenje zank	72
6.9	Izbirni argumenti funkcij in izbirni argumenti funkcije print	75
7	Uporaba in pisanje funkcij	77
7.1	Kaj so funkcije in zakaj so uporabne?	77
7.2	Kako definiramo funkcijo?	78
7.3	Globalni imenski prostor	79
7.4	Kaj se zgodi ob klicu funkcije in lokalni imenski prostor	79
7.5	Vsaka funkcija vrača rezultat	85

7.6	Izbirni argumenti	87
8	Uporaba in pisanje modulov	91
8.1	Kaj so moduli?	91
8.2	Uporaba modulov	91
8.3	Definicija in uporaba lastnih modulov	93
8.4	Nameščanje novih modulov	93
9	Spremenljivost podatkovnih tipov in terke	95
9.1	Kaj je spremenljivost?	95
9.2	Kaj se zgodi ob prirejanju spremenljivk?	96
9.3	Kaj se zgodi ob spreminjanju vrednosti spremenljivk?	96
9.4	Ali funkcije spreminjajo vrednosti svojim argumentom?	98
9.5	Terke	100
9.6	Uporaba terk	100
9.7	Seznami terk in razpakiranje elementov terk	101
9.8	Pakiranje seznamov v seznime terk	102
9.9	Zahteva po nespremenljivosti	104
10	Slovarji	107
10.1	Zakaj slovarji?	107
10.2	Kako uporabljamo slovarje?	108
10.3	Iskanje vrednosti	108
10.4	Dodajanje in spreminjanje vrednosti	109
10.5	Brisanje vrednosti	111
10.6	Ključni in vrednosti	112
10.7	Imenski prostor in slovarji	113
11	Množice	115
11.1	In še množice	115
11.2	Uporaba množic	115
11.3	Omejitve pri uporabi množic	116
11.4	Osnovne operacije nad množicami	116
11.5	Presek, unija in razlika	117
11.6	Metode množic: dodajanje in brisanje elementov	118
11.7	Zgled uporabe množic	119
12	Oblikovanje nizov	127
12.1	Delo z nizi	127
12.2	Iskanje podnizov	129
12.3	Odstranjevanje in spreminjanje (pod)nizov	130
12.4	Razdruževanje in združevanje nizov	131

12.5	Odstranjevanje <i>praznega prostora</i>	132
12.6	Prilagajanje izpisa in formatiranje nizov	134
12.7	Formatiranje nizov in <i>f-Strings</i>	135
13	Delo z datotekami	137
13.1	Zakaj pisati v datoteke in brati iz njih?	137
13.2	Kaj je datoteka?	138
13.3	Tekstovne datoteke	138
13.4	Odpiranje datoteke	139
13.5	Branje datoteke	140
13.6	Pisanje v datoteko	144
13.7	Kodiranje znakov	149
13.8	Datoteke CSV	152
14	Vizualizacija podatkov	157
14.1	Knjižnica Matplotlib in njena namestitve	157
14.2	Funkciji <code>plot</code> in <code>show</code>	157
14.3	Dodajanje oznak	161
14.4	Še malo prilagajanja oznak	162
14.5	Ostale prilagoditve izrisa	166
14.6	Ostali tipi grafov	166
14.7	Risanje matematičnih funkcij	170
15	Knjižnica NumPy in hitro računanje	177
15.1	NumPy in <code>ndarray</code>	177
15.2	Aritmetični operatorji in struktura <code>ndarray</code>	178
15.3	Primerjalni operatorji, indeksiranje s seznamami in filtriranje	178
15.4	Generiranje strukture <code>ndarray</code>	179
15.5	Funkcije nad strukturo <code>ndarray</code>	180
15.6	Več dimenzij	181
15.7	Ostale uporabne funkcije	183
15.8	Posebne vrednosti	184
15.9	Uvažanje vrednosti in omejitve strukture <code>ndarray</code>	184
15.10	Omejitve knjižnice NumPy	188
16	Knjižnica pandas	189
16.1	Delo s podatki	189
16.2	Knjižnica pandas in <code>dataframe</code>	189
16.3	Indeksiranje tabel	191
16.4	Filtriranje vrednosti	193
16.5	Risanje grafov	194
16.6	Izvoz podatkov	197

17 Okolje Jupyter	199
17.1 Interaktivni zvezki	199
17.2 Celice, tipi celic in njihovo poganjanje	199

Osnutek

Predgovor

Pred vami je knjiga, ki je namenjena učenju osnov programiranja v programskem jeziku Python. Kljub temu, da je bila knjiga napisana za študente, ki poslušajo enega izmed predmetov Osnove programiranja na študijih, ki se izvajajo na Fakulteti za kemijo in kemijsko tehnologijo Univerze v Ljubljani, je knjiga primerna za vsakogar, ki bi se rad naučil osnov programiranja. Nekateri zgledi so sicer usmerjeni na področje kemije oziroma biokemije, celotna razlaga pa je splošna in primerna za kogarkoli. Knjiga je predvsem primerna za študente, ki nimajo izkušenj s programiranjem in računalništvo ni njihov osrednji študij. Veliko veselja pri programiranju vam želim!

Miha Moškon, 2020

Osnutek

Zahvala

Zahvalil bi se vsem asistentom in svojim predhodnikom, ki so s svojim delom posredno ali neposredno predmeta Osnove programiranja in Osnove programiranja – izbirni pripeljali do oblike, v kateri se danes izvajata. Prav tako bi se rad zahvalil vsem študentom za njihove pripombe, za vse javljene napake v osnutku knjige ter za vse kritike in pohvale. Hvala!

Osnutek

1 Uvod

1.1 Računalništvo

Računalništvo je v zadnjih letih postalo nepogrešljiv del našega vsakdana. Računalniki niso zgolj naprave, s katerimi igramo igrice, pišemo seminarske naloge in brskamo po internetu, temveč predstavljajo računalniki ključne komponente praktično vseh sodobnih elektronskih ali z elektroniko podprtih naprav od telefonov pa vse do pametnih kuhinjskih aparatov in avtomobilov. Eden od pomembnih dejavnikov, ki je povzročil takšen razmah računalnikov, je zagotovo internet, ki je za novejša generacija zemljanov postal nepogrešljiv učno-izobraževalni, prostočasni in tudi socialni pripomoček.

Računalnike torej srečujemo na vsakem koraku. Prav je, da vsaj približno razumemo, kako računalniki razmišljajo. Če ne drugega, to v današnjem času spada v kontekst splošne razgledanosti. Poleg tega nam razumevanje delovanja računalniških algoritmov odpira tudi malo bolj kritičen pogled do sodobnih tehnologij.

Primer ene izmed povsod prisotnih aplikacij računalništva je umetna inteligenca, ki hote ali nehoti posega v našo zasebnost in preko t.i. priporočilnih sistemov prilagaja zadetke, ki jih dobimo pri iskanju na platformah, kot je Google, izbira oglase, ki bodo prikazani med našim brskanjem po internetu in sortira objave, ki jih vidimo najprej na socialnih omrežjih ter tako (lahko) ustvarja iluzijo sveta, ki nas obdaja. Teorije zarote pravijo, da so se v preteklosti z uporabo manipulacij takih sistemov npr. dobivale predsedniške volitve. Naloga nas kot izobražencev je, da znamo oceniti, ali so take špekulacije upravičene in da se znamo do njih opredeliti in jih kritično ovrednotiti. Dobro je torej, če vsaj približno razumemo kako stvari, ki jih dnevno uporabljamo, delujejo. In to, da se zavedamo, da umetna inteligenca ni čisto nič drugega, kot zbirka različnih matematičnih pristopov, ki je v široko uporabo prešla prav zdaj, ko imamo na voljo veliko količino podatkov (podatki o nas se zbirajo praktično na vsakem našem koraku) in veliko računsko moč (računalniki so v današnjem času dovolj močni, da lahko relativno hitro obdelajo zelo velike količine zajetih podatkov). Dobro je tudi to, da vemo, da lahko odločanje, kaj določenemu uporabniku prikaže t.i. priporočilni sistem, 'nekdo' zlahka zmanipulira, če je motivacija za to dovolj velika.

1.2 Zakaj se učiti programiranja

Dobro je torej vedeti, kako računalniki razmišljajo. Računalniki pa sami po sebi ne razmišljajo, ampak namesto njih 'razmišljajo' programi, ki jih računalniki poganjajo. In programiranje je veja računalništva, ki se ukvarja s pisanjem računalniških programov. Če znamo programirati, torej vsaj približno vemo, na kakšen način računalniki razmišljajo.

Programiranje je poleg tega nadvse uporabno, saj nam omogoča, da pišemo lastne programe in da naloge, ki bi zahtevale, da razmišljamo mi, predamo računalnikom. Obstaja rek, da so vsi dobri programerji veliki lenivci. Obstaja pa tudi prepričanje, da bi moralo učenje programiranja spadati v osnovno izobrazbo vsakega posameznika, saj nas uči algoritmičnega načina razmišljanja. In da bi se morali otroci konceptov programiranja učiti že v prvih letih šolanja, če ne že v vrtcih.

Algoritmičen način razmišljanja se osredotoča na reševanje problemov na tak način, da določimo recept oziroma algoritem, ki opisuje korake, s katerimi bomo problem čimbolj učinkovito rešili. Tak način razmišljanja se zna izkazati za nadvse koristnega pri vsakodnevnem premagovanju problemov. In določanje algoritma predstavlja osnovo pri programiranju. Kompleksen (ali pa tudi ne) problem želimo v tem kontekstu razbiti na sekvenco korakov oziroma sestavin, ki jih bomo uporabili pri reševanju. Te na koncu med seboj povežemo v smiselno zaporedje, to zamešamo v naš program in problem je rešen. Programiranje velikokrat spominja na kuhanje, ko poskušamo določiti recept izbrani jedi, potem pa ta recept prevedemo v mešanje sestavin, njihovo obdelavo z ustreznimi postopki (kuhanje, vzhajanje, peka), včasih pa seveda tudi malo improviziramo. In tako nekako bomo tudi programirali.

1.3 Računalništvo in kemija (in druge vede)

Ne samo da je računalništvo postalo del našega vsakdana, ampak je postalo tudi podporna veda za različna strokovna področja, med katerimi je tudi kemija. Po eni strani računalništvo našemu strokovnemu delu daje podporo in povečuje našo učinkovitost, saj nudi številna orodja za zajem, shranjevanje, obdelavo, vizualizacijo in deljenje podatkov, omogoča krmiljenje kompleksnih naprav itd.

Po drugi strani obstajajo številne aplikacije, ki jih brez računalništva sploh ne bi bilo. Kot prvi primer vzemimo Nobelovo nagrado za kemijo iz leta 2013, ki je bila podeljena za računalniške modele in simulacije v kemiji. Tovrstni modeli odpirajo nova odkritja na področju razumevanja sistemov, ki jih je težko analizirati zgolj eksperimentalno. Drugi primer je uporaba računalnikov pri analizi velike količine eksperimentalnih podatkov, ki jih ročno ne bi mogli nikoli pregledati, kaj šele, da bi iz njih potegnili kaj uporabnega. V tem kontekstu lahko omenimo večletni projekt sekvenciranja in anotacije človeškega genoma, ki je temeljil na podlagi sekvenciranja in sestavljanja fragmentov genoma različnih ljudi. Pri tem je bil projekt zelo

ambiciozen tudi iz računalniške perspektive. Poskrbeti je bilo namreč treba za shranjevanje ogromne količine podatkov in napisati algoritme za odkrivanje vzorcev v podatkih ter združevanje fragmentov v končno celoto. Povezati je bilo potrebno najmodernejše sekvenatorje z najmodernejšimi računalniki oziroma superračunalniki in mediji za shranjevanje velike količine podatkov. Kljub temu, da rezultati uspešno izvedenega projekta niso pripeljali do tako revolucionarnih odkritij, kot so nekateri sprva domnevali (npr. uspešno zdravljenje vseh genskih bolezni), je v sled uspešno zaključenem projektu sekvenciranje genoma posameznika v današnjem času postalo skoraj rutinska klinična metoda za diagnostiko in ocenjevanje tveganja za razvoj določenih bolezni. Določene države so celo investirale sredstva za določitev svojega nacionalnega genoma (na Islandiji so npr. vsaj delno sekvencirali genom več kot polovici prebivalcem) in usmeritev financiranja raziskav v smeri zdravljenja bolezni, za katere so njihovi prebivalci bolj dovzetni. Poleg tega je cenovno dostopno sekvenciranje odprlo nove veje pri razvoju medicine, kot je npr. sistemska in personalizirana medicina, ki na podlagi genskih lastnosti posameznika(ce) poskuša določiti optimalno terapijo posebej zanj(o).

1.4 Kaj je programiranje?

Računalnik lahko v grobem razdelimo na dva dela, in sicer na strojno opremo, ki predstavlja 'platformo', na kateri pač poganjamo programe in ki je brez programske opreme 'mrtva', in programsko opremo, ki predstavlja 'živčni sistem' računalnika.

Programiranje se ukvarja s pisanjem programov oziroma razvojem programske opreme. Proces izvajanja naročil računalniku, tj. naredi nekaj zame, zakomplicira dejstvo, da mi računalnika v osnovi ne razumemo, računalnik pa v osnovi ne razume nas. Programiranje predstavlja proces zapisovanja naših navodil v jeziku, ki je pogojno razumljiv nam (če pač znamo programirati), hkrati pa je razumljiv posebnim programom (ali pa skupini programov), ki naša zaporedja ukazov oziroma kodo, prevedejo (angl. *compile*) ali pretolmačijo (angl. *intepret*) v zapis, ki je razumljiv računalniku. Sekvenco navodil/ukazov oziroma kodo pa lahko pišemo v različnih programskih jezikih. Eni izmed teh so bližje računalnikovem osnovnem jeziku (nižjenivojski jeziki) in nam omogočajo več svobode, imajo pa zato tudi več prostora za to, da lahko naredimo kakšno neumnost ter so na splošno težji za uporabo. V to skupino bi lahko uvrstili npr. programski jezik C. Drugi so bližje našemu načinu razmišljanja (višjenivojski jeziki), zato je programiranje z njimi lažje, je pa izvedba tako napisanih programov mogoče nekoliko počasnejša. V to skupino bi lahko uvrstili npr. programski jezik Python.

1.5 Zakaj Python?

Ta knjiga opisuje programiranje v jeziku Python, ki je po mnenju marsikoga daleč najprimernejši jezik za začetek učenja programiranja. Koda napisana v tem jeziku je enostavno berljiva in intuitivna za razumevanje. Poleg tega programiranje v jeziku Python poteka na nekoliko višjem nivoju (pravimo, da je nivo abstrakcije višji), kar pomeni, da nam ni potrebno zelo natančno vedeti, kaj se npr. v ozadju dogaja s pomnilniškim prostorom, zato lahko s programiranjem brez dodatnega teoretiziranja začnemo kar takoj. Enostavnost jezika pa ne omejuje njegove uporabnosti, Python se namreč uporablja pri razvoju številnih popularnih aplikacij (oziroma programov), kot je npr. brskalnik Google, YouTube, DropBox in Instagram. V letu 2020 je bil tretji najbolj popularen programski jezik (za programskima jezikoma Java in C), njegova popularnost pa še narašča (medtem ko jezikoma Java in C upada).

Poleg tega, da je jezik Python hitro učljiv, to še ne pomeni, da ni zmogljiv. Njegova dodatna prednost je razpoložljivost številnih 'knjižnic',¹ ki razširjajo njegove osnovne funkcionalnosti. Na voljo imamo tudi številne knjižnice, ki jih lahko uporabljamo pri reševanju problemov npr. na področju kemije, kot je knjižnica **chempy** za uporabo v analizi kemiji, knjižnica **chemlab** namenjeno računski kemiji ter vizualizaciji molekul in knjižnica **biopython** za izvajanje bioinformatičnih analiz. Kaj pa počasnost? Hitrost izvajanja programov v jeziku Python je sicer počasnejša od hitrosti programov, ki so napisani v jezikih, kot je C. Vseeno pa lahko pri pisanju programov uporabimo knjižnice, ki 'računsko intenzivne' operacije izvajajo hitreje, ali pa 'računsko intenzivne' dele naših programov napišemo v jeziku, kot je npr. jezik C, in te povežemo z našim Python programom.

Osnov programiranja v jeziku Python se boste lahko torej naučili relativno hitro, poleg tega pa ga boste lahko uporabljali tudi, če se boste po zaključku učenja osnov lotili bolj resnega programiranja.

1.6 Kaj me čaka na koncu knjige?

Ta knjiga je namenjena učenju osnov programiranja v jeziku Python. Pogledali si bomo osnovne gradnike, ki jih uporabljamo pri programiranju in par pogosto uporabljenih knjižnic, predvsem za analizo in vizualizacijo podatkov. Ko boste knjigo prebrali, naj bi znali brez problemov sledeče:

- pisati računalniške programe za zajem podatkov (od uporabnika, iz datotek, iz spleta ipd.),

¹Knjižnica predstavlja zbirko že napisanih delov kode, ki jo lahko enostavno pokličemo iz našega programa. Če bi naredili analogijo s kuhanjem, bi lahko rekli, da imamo z uporabo knjižnic na razpolago bolj kompleksne sestavine, ki jih samo še sestavimo skupaj – npr. namesto, da pečemo testo za torto od začetka, vzamemo že narejen biskvit, ki ga samo še namažemo in okrasimo.

- pisati računalniške programe za obdelavo podatkov (določitev najpomembnejših podatkov, sortiranje podatkov, izvajanje statističnih analiz podatkov),
- pristopiti k vizualizaciji podatkov,
- pristopiti k reševanju problemov na algoritmičen način.

Nenazadnje boste dobili osnovno znanje programiranja, ki ga boste v prihodnosti zlahka nadgradili, če boste tako pač hoteli.

Za konec uvoda pa še opozorilo. Programiranja se še nihče ni naučil z branjem knjig. Kdor hoče znati programirati, mora pač programirati. Knjiga služi zgolj kot pripomoček, ki ga lahko uporabljate pri učenju, njeno branje pa je zagotovo potrebno dopolniti s treningom. Zglede v knjigi torej poskusite v čimvečjem obsegu rešiti sami, rešujte vaje in poskusite znanje programiranja uporabiti pri reševanju problemov, s katerimi se soočate pri drugih predmetih (namesto npr. uporabe orodja Excel). Programiranje je namreč kot šport in za spretnost zahteva svoj trening. Trenirajte.

Osnutek

2 Spoznavanje z okoljem

2.1 Izbira in namestitvev okolja

Razlaga in zgledi v knjigi temeljijo na okolju *Python 3*. Osnovno okolje, ki bo za spoznavanje osnov programiranja čisto dovolj, lahko dobite na Pythonovi domači strani (<https://www.python.org/>), kjer si izberete še svoj operacijski sistem in verzijo, ki jo želite namestiti (izberite si različico Python 3...). Če uporabljate operacijski sistem Linux ali OS X, imate Python po vsej verjetnosti že nameščen. Pazite le na to, da uporabljate Python 3, saj večina zgledov v okolju Python 2 ne bo delovala ali pa bo celo delovala nekoliko drugače. Ob nameščanju okolja Python je koristno, če na začetku obkljukate izbiro `Add Python 3... to PATH`, saj si boste s tem mogoče prihranili kakšno težavo proti koncu knjige. Če ste ambiciozni in pričakujete, da boste Python v prihodnosti bolj resno uporabljali, si mogoče že zdaj namestite distribucijo *Anaconda*, ki vključuje malo več knjižnic za bolj napredno računanje, analizo podatkov in risanje grafov. Dobite jo na strani <https://www.anaconda.com/>.

2.2 Okolje IDLE

Tekom razvoja enostavnejših ali kompleksnih programov ponavadi uporabljamo različna razvojna okolja (angl. *Integrated Development Environment*, IDE), ki združujejo orodja za pisanje, poganjanje in razhroščevanje (angl. *debugging*) programov. Ti pripomočki nam nekoliko olajšajo delo in naredijo razvoj kompleksnejših (ali pa tudi enostavnejših) programov nekoliko bolj pregleden. Z namestitvijo osnovnega okolja Python smo na svoj računalnik namestili tudi okolje IDLE, ki predstavlja enostavno razvojno okolje za jezik Python in katerega bomo uporabljali pri spoznavanju osnov programiranja. Zaženimo ga in začnimo (v operacijskem sistemu Windows boste IDLE zagnali tako, da v Start meni napišete `idle` in pritisnete tipko `Enter`).

2.3 Ukazna vrstica

Pred nami se je pojavila *ukazna* oziroma *pozivna vrstica*, s pomočjo katere lahko začnemo naš pogovor s tolmačem jezika Python (angl. *Python interpreter*). Mi mu bomo v obliki stavkov podajali ukaze, on pa jih bo izvršil in nam ponavadi tudi vrnil ali pa izpisal nek rezultat. Ukaz napišemo in tolmaču pošljemo tako, da pritisnemo tipko **Enter**.

Poskusimo nekaj osnovnih stvari.

```
>>> 1 + 2
3
>>> 1 - 2
-1
>>> 1 * 2
2
>>> 1 - 3 * 4
-11
```

Tolmaču torej lahko podam nek izraz sestavljen iz operandov (v tem primeru števil) in operatorjev (npr. +, * in -) in on mi bo vrnil rezultat. Iz zadnjega izraza vidimo tudi, da ima množenje prednost pred odštevanjem. Uporabim lahko tudi oklepaje, s katerimi določim vrstni red računanja:

```
>>> (1 - 3) * 4
-8
```

Kaj pa deljenje?

```
>>> 5 / 4
1.25
>>> 10 / 5
2.0
```

Uporabim lahko tudi celoštevilsko deljenje (//) in ostanek pri deljenju oziroma operacijo modulo (%):

```
>>> 5 // 4
1
>>> 10 // 5
2.0
>>> 10 % 4
2
```

2.4 Podatkovni tipi

Celoštevilsko deljenje očitno vedno vrne celo število, običajno deljenje pa vrne decimalno število, tudi če je decimalni del tega števila enak 0. Operaciji sta pač definirani tako, da vsakič vrneto določen **podatkovni tip**, ki je poleg tega odvisen tudi od podatkovnega tipa vhodov. Podatkovni tip? Vse vrednosti, ki smo jih tolmaču podali, so imele določen podatkovni tip, in sicer smo v vseh primerih zgoraj podajali cela števila (angl. *integer*), ki jim Python pravi **int**. Rezultati izvedbe gornjih stavkov so v določenih primerih prav tako predstavljali cela števila, pri deljenju pa smo dobili decimalna števila oziroma *števila v plavajoči vejici* (angl. *floating point number*), ki jim Python pravi **float**. Decimalna števila lahko tolmaču podamo tudi kot vhodne podatke:

```
>>> 5.4 + 4.6
10.0
>>> 6.3 / 2
3.15
>>> 4 ** 0.5
2.0
```

Mimogrede, operator ****** določa potenciranje, potenciranje z vrednostjo 0.5 pa je enako kvadratnemu korenu.

Python zna poleg s števili delati tudi z drugimi podatkovnimi tipi, npr. nizi (angl. *string*), oziroma po njegovo s podatkovnim tipom **str**. Nizi vsebujejo sekvenco poljubnih znakov (črke, števila, ločila, posebni simboli ipd.), zapisujemo pa jih znotraj dvojnih (") ali enojnih navednic ('). Poskusimo:

```
>>> "niz1"
'niz1'
>>> 'niz2'
'niz2'
```

Python očitno ne loči med dvojnimi in enojnimi navednicami. Mi jih lahko uporabljamo čisto po navdihu. Kako pa znotraj niza zapisati enojne navednice? Tako da jih ovijemo v dvojne navednice. In obratno. Poskusimo:

```
>>> 'Rekel je: "Daj mi mir!'"
'Rekel je: "Daj mi mir!'"
>>> "'Znamo' programirati."
"'Znamo' programirati."
```

Ali lahko tudi nad nizi izvajamo operacije od prej. Poskusimo:

```
>>> 'niz1' + 'niz2'
'niz1niz2'
>>> 'niz1' + ' ' + 'niz2'
```

```
'niz1 niz2'
>>> 'niz1' * 3
'niz1niz1niz1'
>>> 'niz1' * 'niz2'
TypeError: can't multiply sequence by non-int of type 'str'
```

Python nam je v zadnjem primeru vrnil napako, kar pomeni, da operacija, ki smo jo hoteli izvesti ni podprta. Ko dobimo napako, se je ne ustrašimo, ampak jo preberimo, saj nam sporoča kaj je narobe¹. Python torej množenja dveh nizov med seboj ne podpira. Tudi seštevanje in množenje je nad nizi definirano nekoliko drugače kot nad števili. Kaj pa če niz vsebuje samo števila?

```
>>> '3' + '5'
'35'
>>> '3' * '5'
TypeError: can't multiply sequence by non-int of type 'str'
```

Rezultat je podoben kot prej. Podatkovni tip operanda oziroma podatka torej določa kaj in kako lahko s posameznim podatkom počnemo. Števila lahko npr. seštevamo, če pa poskusimo sešteti dva niza, izvajamo operacijo lepljenja nizov oziroma njihovo *konkatenacijo*.

2.5 Funkcije

Poleg osnovnih operatorjev so v osnovnem okolju jezika Python vgrajene tudi določene *funkcije*. Podobno kot v matematiki tudi pri programiranju funkcije sprejemajo določene vhode oziroma *argumente*. Funkcijo pokličemo tako, da podamo njeno ime in argumente oziroma vhode funkciji zapišemo znotraj oklepajev. Takole:

```
ime_funkcije(argument_1, argument_2, ..., argument_n)
```

Po klicu se bo funkcija izvedla in nekaj koristnega naredila in/ali vrnila nek uporaben rezultat. Če bi želeli izvedeti, kateremu podatkovnemu tipu pripada nek podatek, bi lahko npr. poklicali funkcijo `type`:

```
>>> type(1)
<class 'int'>
>>> type(1.0)
<class 'float'>
>>> type('niz')
<class 'str'>
```

¹Poročila o napakah so v knjigi nekoliko skrajšana in vsebujejo le tip in pojasnilo napake. Python sicer poroča tudi lokaciji napake.

Poglejmo si še funkcijo `abs`, ki izračuna absolutno vrednost podanega argumenta:

```
>>> abs(-1)
1
>>> abs(-1.4)
1.4
>>> abs(5)
5
>>> abs('niz')
TypeError: bad operand type for abs(): 'str'
```

Tudi argumenti funkcij so torej omejeni na določene podatkovne tipe, kar je smiselno. Funkcija `abs` je na primer omejena zgolj na števila, kot so števila tipa `int` ali `float`, saj absolutne vrednosti niza pač ne moremo izračunati.

Python ima vgrajenih še kar nekaj funkcij, ki pa jih bomo večinoma spoznavali sproti.

2.6 Spremenljivke

Do zdaj smo v ukazno vrstico pisali izraze sestavljene in operatorjev in operandov (podatkov). Python je po vsakem pritisku tipke **Enter** podani izraz pognal in vrnil rezultat, ki pa ga je takoj zatem pozabil. Do dobljenega rezultata tako ne moremo več priti drugače, kot da še enkrat podamo enak izraz, ki ga bo Python ponovno ovrednotil in vrnil enak rezultat. Pogosto pa si želimo rezultate izvedenih stavkov zapomniti oziroma želimo, da jih Python shrani za kasneje. Na ta način lahko sestavljamo kompleksnejše izraze (brez da bi pisali dolge kače čez več vrstic), izračunan podatek uporabimo večkrat in z njim delamo različne stvari (npr. uporaba v drugih izrazih, izpis na zaslone, shranjevanje na trdi disk itd.). Python omogoča, da posameznemu podatku dodelimo ime, preko katerega bomo lahko do tega podatka dostopali še kasneje. Takole:

```
>>> x = 1
>>> y = 2 - 3.5
>>> niz = 'abc'
```

Zdaj Python ni ničesar izpisal, je pa vrednost na desni strani *prireditvenega stavka* priredil imenu na levi strani prireditvenega stavka. Izvedli smo torej prireditev vrednosti na desni imenu na levi, ki mu pravimo tudi **spremenljivka**. Pri tem smo uporabili prireditveni operator `=`. Pozor: to ni operator enakosti, saj vedno deluje samo v eno smer, in sicer tistemu, kar napišemo na levi strani, priredi tisto, kar napišemo na desni strani. Če npr. napišemo

```
>>> x = x + 2
```

to ni nerešljiva enačba (kot bi bila v primeru, ko enačaj obravnavamo kot operator enakosti), ampak zgolj pomeni, da vzamemo vrednost, ki stoji za imenom `x`, to vrednost povečamo za 2 in priredimo imenu `x`. Zgornji stavek torej vrednost v spremenljivki `x` poveča za 2. Kako pa dostopamo do vrednosti posamezne spremenljivke oziroma do vrednosti, ki stoji za določenim imenom? To smo naredili že zgoraj – tako da podamo ime spremenljivke. Ko smo zgoraj napisali ime `x` na desni strani prireditvenega stavka, je Python pogledal kaj za tem imenom stoji in ime zamenjal z vrednostjo za njim. Drugače je, če ime uporabimo na levi strani prireditvenega stavka. S tem namreč imenu priredimo novo vrednost, če pa imena pred tem še nismo definirali, s tem ustvarimo tudi novo ime. Temu rečemo **definicija spremenljivke**.

Do vrednosti spremenljivke `x` ali `y` bi zdaj lahko dostopali tako, da podamo njeno ime:

```
>>> x
3
>>> y
-1.5
```

Ime spremenljivke lahko uporabimo tudi kot argument funkcije, npr. takole:

```
>>> abs(x)
3
>>> abs(y)
1.5
```

Prav tako lahko izhod funkcije priredimo novi (ali obstoječi) spremenljivki:

```
>>> z = abs(y)
```

Zdaj Python ni izpisal ničesar, je pa ustvaril novo spremenljivko, do katere lahko dostopamo:

```
>>> z
1.5
```

Kaj pa bi se zgodilo, če pokličemo ime spremenljivke, ki je še nimamo:

```
>>> novo_ime
NameError: name 'novo_ime' is not defined
```

Ker tega imena Python ne pozna, saj ga še nismo definirali, javi napako. Dostopamo lahko torej le do imen, ki smo jih bodisi definirali mi ali pa so že definirana (kot npr. `abs`).

Kaj pa bi se zgodilo, če bi imenu vgrajene funkcije priredili neko vrednost? Če bi npr. izvedli prireditev

```
>>> abs = 5
```

bo Python to brez pritoževanja tudi izvedel. Poskusimo zdaj še enkrat izračunati absolutno vrednost tistega, kar se skriva za spremenljivko `y`:

```
>>> abs(y)
TypeError: 'int' object is not callable
```

Seveda bo prišlo do napake, saj smo si funkcijo za izračun absolutne vrednosti *povozili* z vrednostjo, ki pripada podatkovnemu tipu `int`. Za imenom `abs` po novem Python nima več shranjene funkcije za izračun absolutne vrednosti, ampak število 5. Nerodno. Stvar lahko rešimo tako, da okolje IDLE resetiramo (`Shell` → `Restart Shell` oziroma kombinacija tipk `Ctrl + F6`).

V splošnem velja, da se moramo pri poimenovanju spremenljivk držati določenih pravil. Kot smo videli prej, uporaba imen, ki so že rezervirana oziroma uporabljena, ni priporočena. Prireditve vrednosti rezerviranim imenom ni zgolj slaba, ampak celo vrne napako:

```
>>> if = 5
SyntaxError: invalid syntax
```

Zgoraj vidimo, da je besedica `if` t.i. **rezervirano ime**, saj jo IDLE obarva drugače kot ostale besede oziroma jo odebeli. Njeno uporabo bomo spoznali prav kmalu. Pri imenih spremenljivk nam Python poleg tega ne bo pustil uporabe presledkov:

```
>>> moje ime = 'Miha'
SyntaxError: invalid syntax
```

Stvar lahko rešimo tako, da presledke zamenjamo s podčrtaji (`_`):

```
>>> moje_ime = 'Miha'
>>> moje_ime
'Miha'
```

Pri poimenovanju spremenljivk se je dobro držati določenih priporočil. Navedimo jih nekaj:

- spremenljivke naj imajo smiselna imena, ki programerju sporočajo pomen spremenljivke;
- pri poimenovanju spremenljivk se izogibajmo črkam, ki ne nastopajo v osnovni angleški abecedi (izogibamo se npr. črkam `č`, `ž` in `š`);
- imena so lahko sestavljena iz več besed, pri čemer te ločimo s podčrtaji.

2.7 Pisanje programov

S tolmačem smo se do zdaj pogovarjali preko ukazne vrstice, čemur bi pa težko rekli programiranje. V ukazno vrstico bi lahko sicer napisali zaporedje stavkov,

s katerim bi nekaj izračunali, oziroma s katerim bi rešili nek problem. Če pa bi hoteli to zaporedje stavkov pognati še enkrat (mogoče na drugih podatkih), bi morali v ukazno vrstico stavke v enakem vrstnem redu napisati ponovno. Očitno je, da to ni najbolj priročen način programiranja.

V splošnem zaporedja stavkov zapisujemo v tekstovne datoteke – v programe, ki jih potem v poganjanje predamo našemu tolmaču. Tolmač bo ukaze v datoteki izvedel po vrsti na podoben način, kot če bi te zapisali v ukazno vrstico. Kako lahko za pisanje takih programov uporabimo orodje IDLE? Najprej bomo ustvarili novo datoteko – program, in sicer z izbiro menija **File** → **New File** oziroma s kombinacijo tipk **Ctrl** + **N**. V primeru, da datoteka z nekim programom že obstaja, lahko to odpremo preko menija **File** → **Open** oziroma s kombinacijo tipk **Ctrl** + **O**. V obeh primerih s tem odpremo tudi IDLE-ov urejevalnik teksta, s pomočjo katerega lahko napišemo program, tega shranimo in na koncu poženemo oziroma v poganjanje damo tolmaču. Kot ste verjetno že vajeni, končnica datotek nakazuje kaj približno datoteka vsebuje. Datoteke, v katerih je shranjen program v jeziku Python, prepoznamo preko končnice **.py**. Programe, ki jih bomo pisali, bomo torej tudi mi opremili s tako končnico.

Napišimo krajši program, ki temperaturo v stopinjah Celzija pretvori v temperaturo v stopinjah Fahrenheit z upoštevanjem enačbe $T_F = T_C * 1.8 + 32$, pri čemer T_C predstavlja temperaturo v stopinjah Celzija, T_F pa temperaturo v stopinjah Fahrenheit. Pri tem zaenkrat predpostavljajmo, da je T_C , ki ga želimo pretvoriti, enak 20. Program bo sledeč²:

```
1 T_C = 20
2 T_F = T_C * 1.8 + 32
```

Program lahko zdaj shranimo, npr. v datoteko z imenom **temperatura.py** in poženemo z izbiro menija **Run** → **Run Module** oziroma s pritiskom na tipko **F5**. Izvršil se je preklon na ukazno vrstico, nikjer pa ni vidnega rezultata izvršitve našega programa. Ali se je program res izvedel? Preverimo lahko tako, da v ukazno vrstico napišemo imeni spremenljivk, ki smo jih v programu definirali. Če se program ni pognal, bo Python vrnil napako.

```
>>> T_C
20
>>> T_F
68.0
```

Program se je očitno izvedel, saj sta spremenljivki definirani. Program pa pred tem ni ničesar izpisal. Program ne izpiše ničesar, tudi če ga dopolnimo z zgornjima vrsticama:

²Kadar bomo v zgledih kode vrstice označili z njihovimi številkami, bo to pomenilo, da gre za program.


```

1 T_C = 20
2 T_F = T_C * 1.8 + 32
3 T_C
4 T_F

```

Zakaj ne? Ko svojo kodo zapakiramo v programe, ti izpisujejo vrednosti samo takrat, ko to od njih eksplicitno zahtevamo. Kako? S funkcijo `print`.

2.8 Funkcija `print`

Funkcija `print` nam omogoča izpisovanje vrednosti znotraj programov. Pokličemo jo tako, da ji kot argumente naštejemo vrednosti, ki jih želimo izpisati in funkcija `print` bo vednosti pač izpisala, vmes bo dala presledke, na koncu izpisa pa bo skočila v novo vrstico. Dopolnimo zgornji program, tako da izpiše obe temperaturi:

```

1 T_C = 20
2 T_F = T_C * 1.8 + 32
3 print(T_C)
4 print(T_F)

```

Funkciji smo podali ime spremenljivke, izpisala pa je vrednost, ki stoji za imenom:

```

20
68.0

```

Funkcija `print` je na koncu vsakega izpisa avtomatsko skočila v novo vrstico. Obe temperaturi bi lahko izpisali tudi v isti vrstici, in sicer takole:

```
print(T_C, T_F)
```

V tem primeru je izpis sledeč:

```
20 68.0
```

Funkcija `print` je torej podani vrednosti avtomatsko ločila s presledkom. Poskusimo zdaj izpis še malo olepšati. Poleg imen spremenljivk lahko kot argumenti nastopajo tudi konstantne vrednosti. Lahko bi v izpis dodali še niz, ki bi naredil vse skupaj malo bolj informativno. Takole

```
print(T_C, "°C je enako", T_F, "°F.")
```

V tem primeru je izpis sledeč:

```
20 °C je enako 68.0 °F.
```

Imena spremenljivk je Python torej zamenjal z njihovimi vrednostmi, nize pa je izpisal kakor so bili pač podani.

2.9 Funkcija `input`

Še malo pa bomo napisal naš prvi pravi program. Pretvarjanje iz stopinje Celzija v stopinje Fahrenheit sicer deluje, malo pa je moteče to, da lahko pretvarjamo samo eno vrednost, ki je že vnaprej določena. Program bi bil bistveno bolj uporaben, če bi lahko vrednost, ki jo želimo pretvoriti, podal kar uporabnik ob zagonu programa. To nam omogoča funkcija `input`.

Funkcija `input` prav tako kot `print` na zaslon izpiše podan argument. Za razliko od funkcije `print`, funkcija `input` sprejema samo en argument tipa `str`. Preko tega argumenta bomo funkciji `input` podali navodilo za uporabnika. Na primer takole:

```
>>> input("Vnesi svoje ime: ")
```

Funkcija `input` čaka na uporabnikov vnos in pritisk tipke `Enter`. Če smo to funkcijo pognali iz ukazne vrstice, bo uporabnikov vnos ponovila. Lahko pa uporabnikov vnos shranimo v spremenljivko in ga kasneje uporabimo. Takole

```
>>> ime = input("Vnesi svoje ime: ")
```

Rezultat funkcije smo torej priredili spremenljivki `ime`. Zdaj lahko do tistega, kar je uporabnik vnesel, dostopamo preko imena spremenljivke. Uporabimo to na našem programu za pretvarjanje med temperaturami. Temperaturo v stopinjah bomo zdaj prebrali od uporabnika preko funkcije `input`.

```
1 T_C = input("Vnesi temperaturo v °C: ")
2 T_F = T_C * 1.8 + 32
3 print(T_C, "°C je enako", T_F, "°F.")
```

Program pa v tej obliki žal še ne bo deloval. Funkcija `input` namreč vedno vrne niz, saj je to podatkovni tip, v katerega lahko zapiše karkoli bo pač uporabnik vnesel. Tudi če bo uporabnik vnesel število, bo to predstavljeno kot niz oziroma podatkovni tip `str`. Kakor se spomnimo od prej pa nizov ne moremo množiti z decimalnimi števili pa tudi seštevanje je denifinirano tokrat za nas nekoliko neugodno. Prebrano število, ki je zapisano kot niz, moramo torej pred nadaljnjo obdelavo pretvoriti v nekaj, s čimer lahko računamo, npr. `float`.

2.10 Pretvarjanje med podatkovnimi tipi

Pretvorbo podatka v posamezen podatkovni tip lahko izvedemo z vgrajenimi funkcijami, ki nosijo enako ime, kot podatkovni tip, v katerega želimo pretvarjati. Če bi želeli nekaj pretvoriti v niz, bi torej uporabili funkcijo `str`, v celo število funkcijo `int`, v decimalno pa funkcijo `float`. Poskusimo:

```
>>> niz = str(20)
>>> niz
```

```
'20'
>>> int(niz)
20
>>> float(niz)
20.0
>>> x = 5.4
>>> str(x)
'5.4'
>>> int(x)
5
```

Rezultat izvedbe posamezne funkcije je torej zapis podatka v željenem podatkovnem tipu. Pri tem so seveda upoštevane omejitve posameznega podatkovnega tipa. Ko npr. pretvarjamo v podatkovni tip `int`, funkcija `int` poreže decimalke za decimalno piko (brez zaokroževanja).

```
>>> int(6.9)
6
```

Drugi primer take omejitve je, da nize lahko pretvarjamo v števila le, ko vsebujejo zgolj in samo števila (in decimalno piko):

```
>>> float("6.9")
6.9
>>> int("6a")
ValueError: invalid literal for int() with base 10: '6a'
>>> float("stevilo")
ValueError: could not convert string to float: 'stevilo'
```

Dokončajmo zdaj naš prvi program.

Zgled 1 *Napiši program, ki od uporabnika prebere temperaturo v stopinjah Fahrenheit in to pretvori v stopinje Celzija ter poda izpis obeh temperatur.*

Rešitev 1 *Rešitev naloge praktično že imamo. Malenkost jo moramo le še dopolniti.*

```
1 niz = input("Vnesi temperaturo v °C: ")
2 T_C = float(niz)
3 T_F = T_C * 1.8 + 32
4 print(T_C, "°C je enako", T_F, "°F.")
```

Vrstici številka 1 in 2 bi lahko združili tudi v eno samo:

```
1 T_C = float(input("Vnesi temperaturo v °C: "))
2 T_F = T_C * 1.8 + 32
3 print(T_C, "°C je enako", T_F, "°F.")
```

Program zdaj deluje kakor bi želeli. Do problema pride samo takrat, ko uporabnik vnese kakšno neumnost.

2.11 Pisanje komentarjev

Ponavadi je dobro, da programe pišemo na tak način, da jih bodo razumeli tudi drugi in da jih bomo mogoče razumeli sami, ko jih bomo čez par mesecev ponovno pregledovali. Pri tem pomaga že to, da se trudimo pisati lepo in pregledno kodo ter spremenljivke poimenujemo tako, da vsaj približno vemo kaj predstavljajo (`T_F` in `T_C`). Dodatno pa k razumevanju napisanih programov pripomorejo komentarji. Komentarji služijo opombam, ki jih sebi ali drugim pišemo znotraj naših programom, niso pa namenjeni izvajanju. Pythonu moramo torej nekako povedati, da naj komentarjev ne poganja. To mu lahko sporočimo z uporabo določenih znakov. Znak `#` označuje vrstični komentar, in sicer bo Python preskočil vso kodo od začetka znaka `#` do konca vrstice, v kateri se ta znak nahaja. Primer uporabe je sledeč:

```
>>> # napišemo lahko karkoli
>>> x = 1 # prireditve bo pognana, komentar pa ne
>>> x
1
```

Včasih si želimo napisati daljši (večvrstični komentar). Tega začnemo s tremi enojnimi ali tremi dvojnimi narekovaji in ga končamo s tremi enojnimi ali tremi dvojnimi narekovaji (s tistimi pač, s katerimi smo komentar začeli).

Povadimo zdaj oboje na našem prvem programu.

```
1  """
2  Program, ki pretvori prebrano temperaturo
3  iz stopinj Celzija v stopinje Fahrenheit
4  """
5
6  # branje in pretvorba v decimalno število
7  T_C = float(input("Vnesi temperaturo v °C: "))
8
9  T_F = T_C * 1.8 + 32 # enačba pretvorbe
10
11 # izpis na zaslon
12 print(T_C, "°C je enako", T_F, "°F.")
```

3 Pogojni stavek

3.1 Zakaj pogojni stavki?

Vsi programi, ki smo jih do zdaj napisali (je bil mogoče samo eden?), so se izvedli po vnaprej določenem zaporedju. Od zgoraj navzdol so se namreč lepo po vrsti izvedli vsi stavki v programu. V določenih primerih pa bi posamezne stavke radi izvedli samo ob izpolnjenosti (ali pa neizpolnjenosti) izbranega pogoja. Poglejmo si spodnji primer:

Zgled 2 *Napiši program, ki od uporabnika prebere telesno maso in višino in izpiše uporabnikov indeks telesne mase.*

Rešitev 2 *Preko funkcije `input` bomo torej prebrali telesno maso in višino. Ker funkcija vrača niz, bomo obe vrednosti pretvorili v decimalno število (`float`). Potem bomo uporabili enačbo za izračun indeksa telesne mase: $itm = \frac{masa}{visina^2}$. Pri tem mora biti telesna masa podana v kilogramih, višina pa v metrih.*

```
1 masa = float(input("Vpiši svojo telesno maso [kg]: "))
2 visina = float(input("Vpiši svojo višino [m]: "))
3 itm = masa/visina**2
4 print("Tvoj ITM je", itm)
```

Zgornji program je popolnoma pravilen, bi ga pa radi še malo dopolnili. Veliko uporabnikov verjetno ne ve kaj posamezna vrednost indeksa telesne mase (ITM) pomeni. Ali mora shujšati, je njegova telesna masa ustrezna, ali bi se moral malo zrediti? Če malo poenostavimo, lahko ljudi razdelimo v tri skupine, in sicer na tiste s premajhno telesno maso, tiste z ustrezno telesno maso in tiste s preveliko telesno maso:

pogoj	skupina
$ITM < 18.5$	podhranjenost
$18.5 \leq ITM \leq 25$	normalna telesna masa
$25 < ITM$	debelost

Naš program bi torej radi dopolnili tako, da bo uporabniku izpisal tudi informacijo o tem, v katero skupino spada. Z drugimi besedami, če bo izpolnjen prvi pogoj ($ITM < 18.5$), bi radi izpisali, da je uporabnik podhranjen, če bo izpolnjen drugi

pogoj ($18.5 \leq \text{ITM} \leq 25$), da je njegova telesna masa ustrezna in če bo izpolnjen tretji pogoj ($25 < \text{ITM}$), da je pretežak. Radi bi torej, da se določeni deli našega programa (v konkretnem primeru različni izpisi) izvedejo v odvisnosti od vrednosti ITM.

3.2 Osnovna oblika stavka `if`

Za pisanje pogojnih stavkov bomo uporabili Pythonov stavek `if`. Njegova osnovna oblika je sledeča:

```
if pogoj :
    # pogojni_stavki so zamaknjeni
    # pogojni stavki
    # če je pogoj izpolnjen
    ...
# skupni stavki
# ni več zamaknjeno
...
```

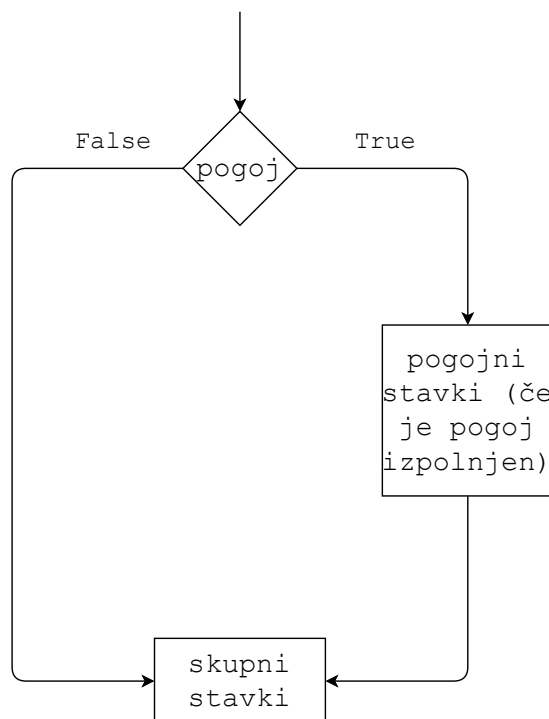
Začnemo torej z rezervirano besedico `if`, ki ji sledi pogoj. Temu sledi dvopičje (`:`), s katerim povemo, da je konec pogoja. Potem sledi pogojni del, ki se bo izvedel samo v primeru, da je podan pogoj. Pogojnemu delu sledijo stavki, ki se bodo izvedli v vsakem primeru, ne glede na izpolnjenost pogoja. Kako Pythonu povemo kje je konec pogojnega dela? Z zamikom (angl. *indent*). Zgoraj smo tiste stavke, ki se izvedejo samo v primeru izpolnjenosti pogoja zamaknili, tako da smo pred njih vstavili tabulator (tipka *Tab*) ali par presledkov (če smo natančni, se držimo dogovora štirih presledkov). Pogojni del smo zaključili tako, da smo enostavno nehali zamikati. Izvedbo zgornje kode prikazuje diagram poteka na sliki 3.1.

3.3 Kaj je pogoj?

Kaj pravzaprav predstavlja pogoj za izvedbo pogojnega dela stavka? Kot je razvido iz slike 3.1 je pogoj nekaj, kar je lahko resnično (angl. *true*) ali neresnično (angl. *false*). Pogoj moramo torej zastaviti kot vprašanje, na katerega lahko odgovorimo bodisi z odgovorom *da* ali z odgovorom *ne*. Pri formiranju vprašanja oziroma pogoja bomo torej večinoma uporabljali operatorje, ki vračajo take odgovore. Tem operatorjem pravimo *pogojni operatorji* pa si pogledjmo nekaj njihovih primerov.

3.3.1 Primerjalni operatorji in podatkovni tip `bool`

Prva skupina operatorjev, ki jih bomo uporabljali pri pisanju pogojev so t.i. *primerjalni operatorji*, ki jih poznamo že iz matematike. To so npr. operator enakosti



Slika 3.1 Diagram poteka osnovne oblike stavka `if`. V primeru, če je pogoj izpolnjen, se izvede pogojni del.

`==` (ker je enojni enačaj uporabljen za prirejanje, moramo za primerjanje uporabiti dvojni enačaj), operator neenakosti `!=`, večji `>`, večji ali enak `>=` itd. Poskusimo:

```

>>> 1 == 1
True
>>> 1 == 2
False
>>> 1 != 2
True
>>> 1 > 2
False
>>> 1 <= 2
True
>>> 1 == 2.5
False
>>> 1 == 1.0
True

```

S primerjalnimi operatorji torej med seboj primerjamo dva podatka, rezultat primerjanja pa je bodisi vrednost `True` ali `False`. Rezultat primerjanja je podatek,

ki lahko zavzame samo te dve vrednosti. Kakšen je podatkovni tip tega podatka?

```
>>> type(True)
<class 'bool'>
>>> type(False)
<class 'bool'>
```

Za oblikovanje pogojev imamo torej na voljo poseben podatkovni tip, tj. `bool`, oziroma po angleško *boolean*, ki lahko zavzame samo dve vrednosti, tj. `True` ali `False`.

Poskusimo uporabo primerjalnih operatorjev v kombinaciji s pogojnimi stavki `if` uporabiti na zgledu iz začetka poglavja.

Zgled 3 *Napiši program, ki od uporabnika prebere telesno maso in višino in izpiše uporabnikov indeks telesne mase (ITM). Poleg tega program uporabniku pove, v katero skupino spada.*

Rešitev 3 Program od prej bomo dopolnili s pogojnimi stavkami. Če je ITM manjši od 17.5, lahko program izpiše, da je uporabnikova telesna masa premajhna. Če je ITM večji od 25, lahko program izpiše, da je uporabnikova telesna masa prevelika. Kaj pa vmes? Tega pa zaenkrat še ne znamo.

```
1 masa = float(input("Vpiši svojo telesno maso [kg]: "))
2 visina = float(input("Vpiši svojo višino [m]: "))
3 itm = masa/visina**2
4 print("Tvoj ITM je", itm)
5 if ITM < 17.5:
6     print("Tvoja telesna masa je premajhna")
7 if ITM > 25:
8     print("Tvoja telesna masa je prevelika")
```

Primerjalne operatorje lahko uporabimo tudi nad podatki, ki niso števila

```
>>> "abc" == "abc"
True
>>> "abc" == "ABC"
False
>>> "abc" < "b"
True
>>> "abc" < "abc"
False
>>> "abc" < "abd"
True
```


Iz zgornjih zgledov vidimo tudi to, da Python loči med velikimi in malimi črkami in da so določeni nizi manjši od drugih. Kako pa primerjanje dveh nizov poteka? Na enak način, kot primerjamo nize, ko poskušamo besede sortirati po abecedi (npr. v slovarju ali telefonskem imeniku). Dve besedi primerjamo znak po znaku od začetka do konca, dokler ne pridemo do dveh znakov, ki se razlikujeta ali do konca ene izmed besed. Če se besedi ujemata po vseh znakih in je ena beseda krajša, je krajša beseda zagotovo manjša. Npr., beseda "beda" je manjša od besede "bedarija" (v slovarju bo verjetno beda nastopala pred bedarijo):

```
>>> "beda" < "bedarija"
True
```

Beseda "bedno" pa ni manjša od besede "bedarija", čeprav je od nje krajša. Zakaj ne? Zato, ker se besedi razlikujeta v četrtem primerjanju na znakih "n" in "a" in ker "n" pač ni manjši od znaka "a".

```
>>> "bedno" < "bedarija"
False
```

Takemu primerjanju pravimo *leksikografsko* primerjanje.

3.3.2 Operatorja vsebovanosti

Ko smo ravno pri nizih, lahko omenimo še *operatorja vsebovanosti*, ki preverjata ali nekaj je (*in*) oziroma ni (*not in*) v posameznem nizu vsebovano. Operatorja bomo uporabljali tudi na drugih podatkovnih tipih, ki podobno kot nizi, vsebujejo druge podatke – nizi so sestavljeni iz več znakov oziroma podnizov. Če je nek niz podniz vsebovan v nekem nizu niz, lahko preverim takole:

```
>>> podniz in niz
```

Povadimo:

```
>>> "beda" in "bedarija"
True
>>> "Beda" in "bedarija"
False
>>> "ana" in "anakonda"
True
>>> "a" in "abeceda"
True
```

Spet vidimo, da je znak "b" nekaj drugega kot znak "B" in da Python loči med malimi in velikimi črkami.

3.3.3 Združevanje rezultatov primerjanja

Pri reševanju naloge z izpisovanjem podatkov o ITM imamo še vedno težave s primerom, kjer morata biti izpolnjena dva pogoja hkrati ($ITM \geq 18.5$ in $ITM \leq 25$). Končen pogoj za izvedbo izpisa *Tvoja telesna masa je ustrezna*, moramo torej sestaviti iz dveh pogojev. Za ta namen lahko uporabimo t.i. *logične operatorje*, ki omogočajo medsebojno združevanje več spremenljivk tipa `bool`. Osnovna logična operatorja, ki ju bomo uporabljali v takem primeru sta operator `and` in operator `or`. Njuno delovanje lahko ponazorimo s spodnjo tabelo:

pogoj1	pogoj2	pogoj1 and pogoj2	pogoj1 or pogoj2
False	False	False	False
False	True	False	True
True	False	False	True
True	True	True	True

V primeru, da morata biti izpolnjena oba vhodna pogoja, torej uporabimo operator `and`. Če je dovolj, da je izpolnjen samo eden izmed vhodnih pogojev, uporabimo operator `or`. Pogosto uporabljen logični operator je še operator `not`, ki `True` spremeni v `False` in obratno:

```
>>> not True
False
>>> not False
True
```

Zdaj lahko dokončamo zgled z izpisovanjem podatkov o ITM.

Zgled 4 *Napiši program, ki od uporabnika prebere telesno maso in višino in izpiše uporabnikov indeks telesne mase (ITM). Poleg tega program uporabniku pove, v katero skupino spada.*

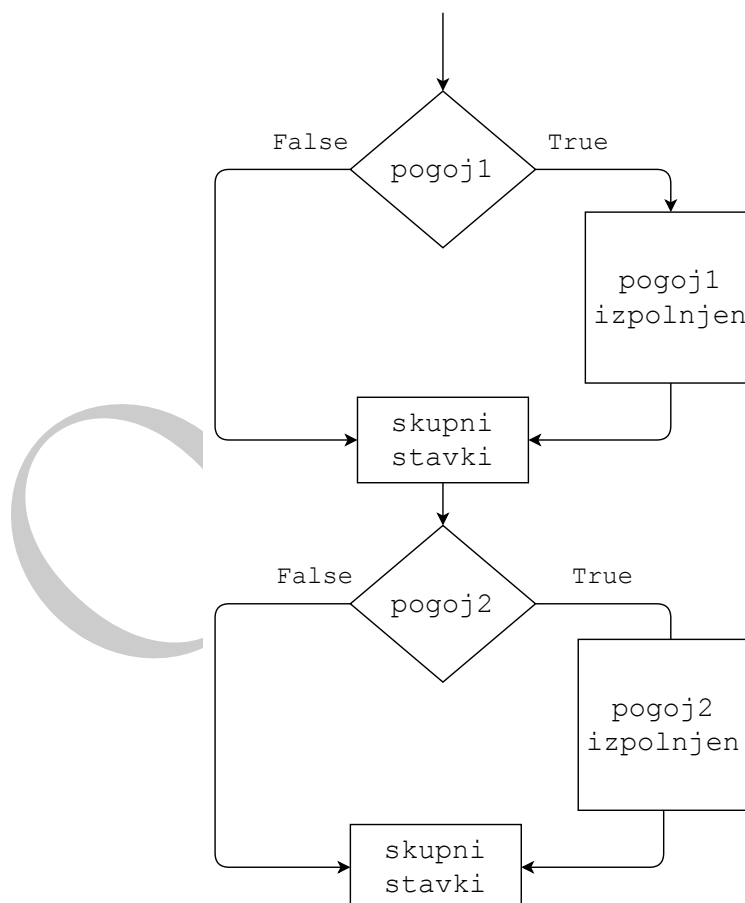
Rešitev 4 *Zdaj lahko dodamo še pogojni stavek, pri katerem bo pogoj sestavljen iz dveh delov. V tem primeru mora biti vrednost spremenljivke *ITM* večja ali enaka od 17.5 in manjša ali enaka od 25, kar lahko zapišemo s pogojem $17.5 \leq ITM$ and $ITM \leq 25$.*

```
1 masa = float(input("Vpiši svojo telesno maso [kg]: "))
2 visina = float(input("Vpiši svojo višino [m]: "))
3 itm = masa/visina**2
4 print("Tvoj ITM je", itm)
5 if ITM < 17.5:
6     print("Tvoja telesna masa je premajhna")
7 if ITM > 25:
8     print("Tvoja telesna masa je prevelika")
9 if 17.5 <= ITM and ITM <= 25:
10    print("Tvoja telesna masa je ustrezna")
```

Programiranja se učimo v jeziku Python med drugim tudi zato, ker ima kar nekaj uporabnih sladkorčkov (funkcionalnosti), ki jih drugi jeziki nimajo. Sestavljen pogoj $17.5 \leq \text{ITM} \text{ and } \text{ITM} \leq 25$ lahko v tem jeziku zapišemo tudi malo krajše, in sicer takole: $17.5 \leq \text{ITM} \leq 25$.

3.4 Veja else

Zgornji program je sicer pravilen, ni pa najlepši. V primeru, da je npr. izpolnjen prvi pogoj, tj. $\text{ITM} < 17.5$, ni nobene potrebe po tem, da preverjamo še izpolnjenost drugega in tretjega pogoja. To sicer v tem primeru ni narobe (lahko bi bilo), je pa nepotrebno in po eni strani naredi našo kodo manj pregledno, po drugi strani pa trati dragocen procesorski čas, saj preverja, če je določen pogoj izpolnjen, kljub temu, da vemo, da zagotovo ne more biti. Potek programa, ki smo ga napisali zgoraj, ponazarja slika 3.2.

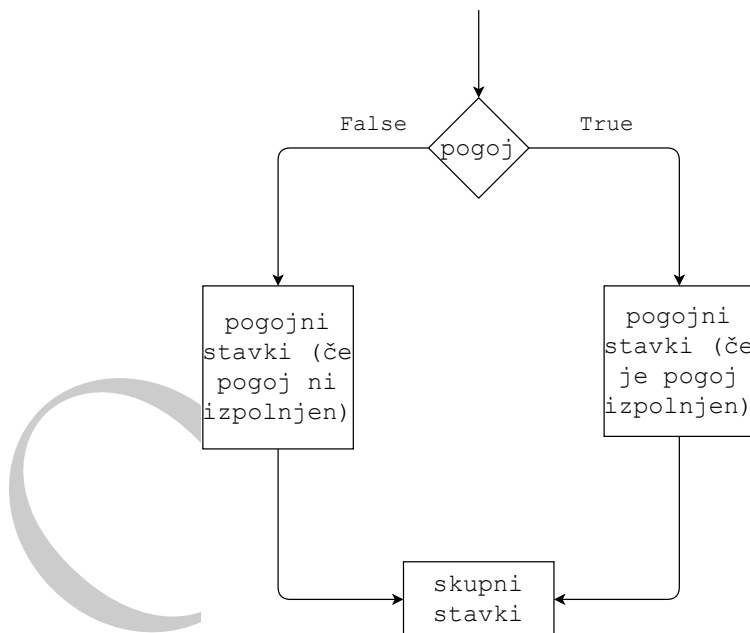


Slika 3.2 Izpolnjenost pogoja `pogoj2` se preverja ne glede na izpolnjenost pogoja `pogoj1`.

Do sedaj smo v primeru neizpolnjenosti pogoja vedno skočili na del **skupni stavki**, torej na del, ki se izvede neodvisno od izpolnjenosti pogoja. V splošnem pa stavek **if** omogoča, da del kode izvedemo samo takrat, ko pogoj **ni** izpolnjen. To kodo podamo v veji **else** stavka **if**:

```
if pogoj:
    # pogojni stavki
    # če je pogoj izpolnjen
    ...
else:
    # pogojni stavki
    # če pogoj ni izpolnjen
    ...
# skupni stavki
...
```

Potek izvajanja zgornje kode prikazuje slika 3.3.



Slika 3.3 Dopolnitev stavka **if** z vejo **else**. Veja **else** se izvede samo v primeru, ko pogoj ni izpolnjen.

Uporabimo zgornji stavek za poenostavljeno rešitev naloge z izpisovanjem podatkov o ITM.

Zgled 5 *Napiši program, ki od uporabnika prebere telesno maso in višino in izpiše uporabnikov indeks telesne mase (ITM). Poleg tega program uporabniku pove, če je*

njegova telesna masa ustreza ali ne.

Rešitev 5 *Tokrat bomo preverjali le pogoj o ustreznosti uporabnikove telesne mase.*

```
1 masa = float(input("Vpiši svojo telesno maso [kg]: "))
2 visina = float(input("Vpiši svojo višino [m]: "))
3 itm = masa/visina**2
4 print("Tvoj ITM je", itm)
5 if 17.5 <= ITM <= 25:
6     print("Tvoja telesna masa je ustreza")
7 else:
8     print("Tvoja telesna masa ni ustreza")
```

3.5 Veja `elif` in gnezdenje stavkov `if`

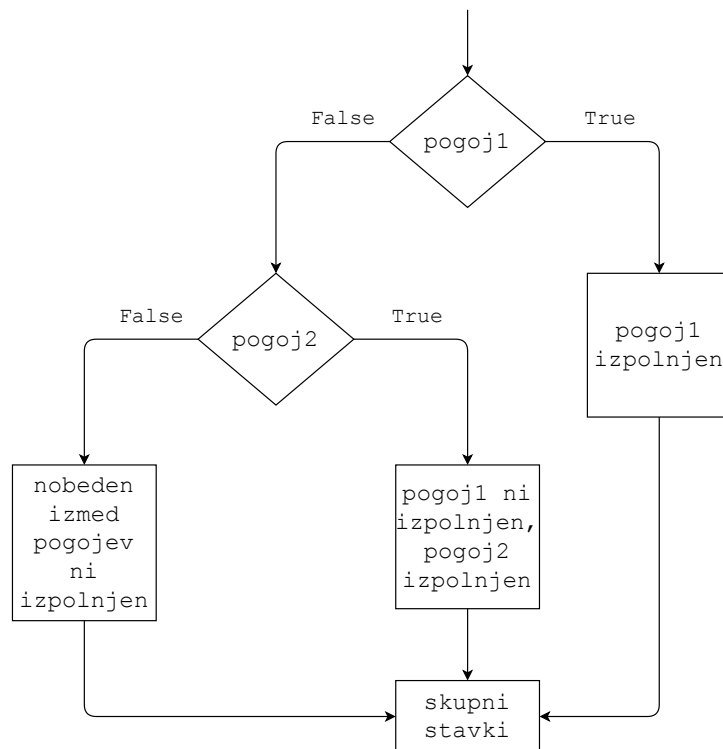
Uporabnik zdaj ve, če je njegova telesna masa ustreza. Če njegova telesna masa ni ustreza, informacije o tem ali je pretežak ali prelahk nima (verjetno se mu to sicer dozdeva). Zgornji primer bi torej radi dopolnili tako, da znotraj veje `else` izvedemo dodatno primerjanje, na podlagi katerega bomo lahko ugotovili ali je ITM prevelik ali premajhen.

To lahko naredimo na dva načina. Elegantnejši način je uporaba stavka `elif`, ki omogoča preverjanje dodatnega pogoja znotraj veje `else`. Celoten stavek `if` z vejo `elif` zapišemo takole:

```
if pogoj1:
    # pogojni stavki
    # pogoj1 izpolnjen
    ...
elif pogoj2:
    # pogojni stavki
    # pogoj1 ni izpolnjen
    # pogoj2 izpolnjen
    ...
else:
    # nobeden izmed pogojev
    # ni izpolnjen
# skupni stavki
...
```

V tem primeru se izpolnjenost pogoja `pogoj2` preverja samo, če pogoj `pogoj1` ni izpolnjen, veja `else` pa se izvede samo v primeru, ko ni bil izpolnjen nobeden izmed prejšnjih pogojev. Potek programa prikazuje slika 3.4.

Zdaj lahko končno podamo lepšo rešitev zгледа z izpisovanjem podatkov o ITM.



Slika 3.4 Dopolnitev stavka `if` z vejama `elif` in `else`. Veja `elif` se izvede samo v primeru, ko pogoj `pogoj1` ni izpolnjen, pogoj `pogoj2` pa je.

Zgled 6 *Napiši program, ki od uporabnika prebere telesno maso in višino in izpiše uporabnikov indeks telesne mase (ITM). Poleg tega program uporabniku pove, v katero skupino spada.*

Rešitev 6 *Najprej bomo preverili enega izmed pogojev, npr. če je ITM manjši od 17.5. V primeru, da je pogoj izpolnjen, izpišemo, da je telesna masa premajhna. V veji `elif` lahko preverimo naslednji pogoj, npr. če je ITM večji od 25. V primeru, da je izpolnjen ta pogoj, izpišemo, da je telesna masa prevelika. Če ni izpolnjen nobeden izmed obeh pogojev, lahko izpišemo, da je telesna masa ustrezna.*

```

1 masa = float(input("Vpiši svojo telesna masa [kg]: "))
2 visina = float(input("Vpiši svojo višino [m]: "))
3 itm = masa/visina**2
4 print("Tvoj ITM je", itm)
5 if ITM < 17.5:
6     print("Tvoja telesna masa je premajhna")
7 elif ITM > 25:
8     print("Tvoja telesna masa je prevelika")
9 else:

```

```
10      print("Tvoja telesna masa je ustrezna")
```

Drug pristop k reševanju enakega problema je uporaba dodatnega stavka `if` znotraj veje `else`. Temu rečemo tudi *gnezdenje* ali *ugnezdeni stavek if*. Kodo bi napisali takole:

```
if pogoj1:
    # pogojni stavki
    # pogoj1 izpolnjen
    ...
else
    if pogoj2:
        # tukaj so uporabljeni dvojni zamiki
        # pogojni stavki
        # pogoj1 ni izpolnjen
        # pogoj2 izpolnjen
        ...
    else:
        # tukaj so uporabljeni dvojni zamiki
        # nobeden izmed pogojev
        # ni izpolnjen
# skupni stavki
...
```

Začetek ugnezdenega stavka `if` je zamaknjen enkrat, s čimer povemo, da naj se izvede samo v primeru, ko pogoj `pogoj1` ni izpolnjen. Vsebino ugnezdenega stavka moramo zamakniti dvakrat. Izvedba zgornje kode bo enaka kot v primeru z uporabo veje `elif` in jo prikazuje slika 3.4.

Uporabimo ugnezden stavek `if` še pri reševanju naše naloge.

Zgled 7 *Napiši program, ki od uporabnika prebere telesno maso in višino in izpiše uporabnikov indeks telesne mase (ITM). Poleg tega program uporabniku pove, v katero skupino spada. Uporabi ugnezden stavek `if`*

Rešitev 7 *Potek programa bo podoben kot prej za razliko od gnezdenja stavka `if`.*

```
1 masa = float(input("Vpiši svojo telesno maso [kg]: "))
2 visina = float(input("Vpiši svojo višino [m]: "))
3 itm = masa/visina**2
4 print("Tvoj ITM je", itm)
5 if ITM < 17.5:
6     print("Tvoja telesna masa je premajhna")
7 else
8     if ITM > 25:
```

```
9         print("Tvoja telesna masa je prevelika")
10    else:
11        print("Tvoja telesna masa je ustrezna")
```

Osnutek

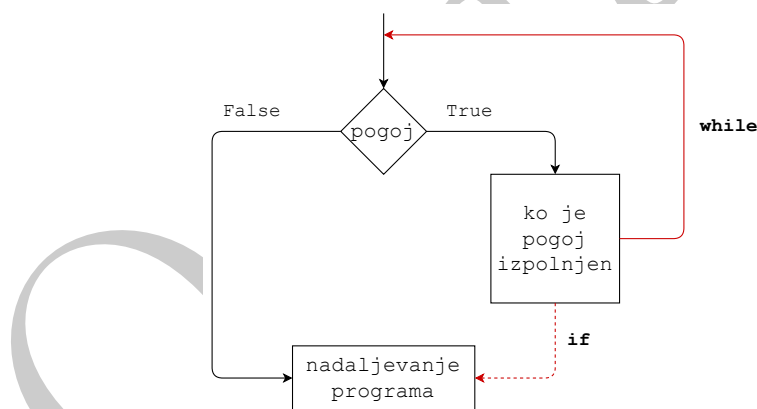
4 Zanka while

4.1 Kaj so zanke?

Z uporabo stavka `if` lahko torej izbrane stavke izvedemo samo v primeru, ko je nek pogoj izpolnjen. Včasih pa bi želeli izbrane stavke izvajati vse **dokler** (angl. *while*) je nek pogoj izpolnjen. To omogočajo **zanke**. V sledečem poglavju si bomo podrobneje pogledali zanko `while`.

4.2 Zanka while

Razliko med izvedbo pogojnega stavka `if` in zanko `while` prikazuje slika 4.1.



Slika 4.1 Razlika med izvedbo pogojnega stavka `if` (črtkana linija rdeče barve) in zanko `while` (polna linija rdeče barve).

Po izvedbi pogojnega dela stavka `if` se izvajanje programa nadaljuje v delu, ki sledi pogojnemu stavku. Po drugi strani se po izvedbi pogojnega dela zanke `while`, recimo tem stavkom raje *telo zanke*, izpolnjenost pogoja v *glavi zanke* ponovno preveri. Telo (vsebina) zanke se bo torej izvajalo vse dokler bo pogoj izpolnjen. Zanko `while` lahko zapišemo takole:

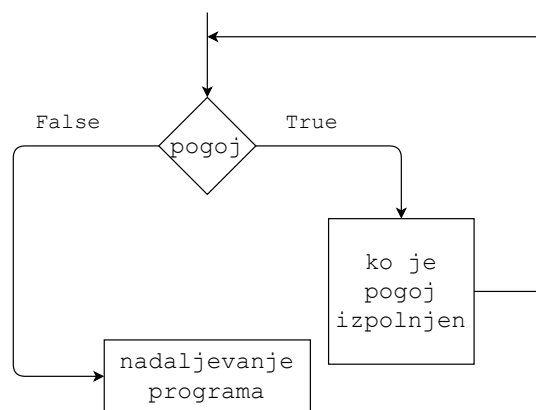
```
while pogoj: # glava zanke
```

```

    # telo zanke
    ...
# nadaljevanje programa
...

```

Zapis zanke `while` je torej zelo podoben zapisu stavka `if`. Glavi zanke sledi telo zanke oziroma njena vsebina, ki se izvaja vse dokler je pogoj izpolnjen. Enemu obhodu zanke pravimo tudi *iteracija* zanke. Pogoje za izvedbo nove iteracije zanke lahko zapisujemo na popolnoma enak način kot pri stavku `if`. Prav tako kot pri stavku `if` vsebino zanke definiramo tako, da stavke znotraj telesa zanke zamikamo. Izvajanje zanke `while` ponazarja diagram poteka na sliki 4.2.



Slika 4.2 Potek izvajanja zanke `while`.

Zanke torej lahko uporabljamo takrat, ko želimo nekaj ponavljati, dokler je določen pogoj izpolnjen. Npr., dokler uporabnik ne poda veljavnega vnosa, dokler ni spremenljivka, ki deluje kot števec, dosegla določene vrednosti ali pa dokler sta števili različni. Uporabo zanke `while` lahko povadimo na programu, ki poišče največji skupni delitelj dveh števil.

Zgled 8 *Napiši program, ki od uporabnika prebere dve celi števili in poišče največji skupni delitelj teh dveh števil z uporabo Evklidovega algoritma.*

Rešitev 8 *Osnovna različica Evklidovega algoritma deluje tako, da manjše število odšteva od večjega, dokler sta števili različni. Ko števili postaneta enaki, smo našli skupnega delitelja. Program bo torej manjše število odšteval od večjega, dokler sta števili različni. Uporabili bomo zanko `while` (dokler sta števili različni). Znotraj zanke bomo uporabili še stavek `if`, s pomočjo katerega bomo ugotovili, katero število je manjše. Ko bosta števili postali enaki, bomo enega izmed njih izpisali (vseeno katerega, ker sta enaki), saj ta predstavlja največji skupni delitelj števil, ki sta bili podani na začetku.*

```

1 st1 = int(input("Vnesi prvo število: "))
2 st2 = int(input("Vnesi drugo število: "))
3
4 while st1 != st2: # dokler sta števili različni
5     if st1 < st2: # drugo število je večje
6         st2 = st2 - st1
7     else: # prvo število je večje
8         st1 = st1 - st2
9
10 # konec vsebine zanke
11 # števili sta tu enaki, zato je vseeno katero izpišem
12 print(st1)

```

Opomba: vsebina stavka if je zamaknjena dvakrat, saj je zapisana tako znotraj stavka if kot tudi znotraj zanke while!

4.3 Štetje z zanko while

Zanko while bi lahko uporabili tudi za štetje. Za ta namen je sicer boljša zanka for, ki jo bomo spoznali malo kasneje. Poglejmo si zgled.

Zgled 9 *Napiši program, ki šteje od 0 do števila, ki ga je vnesel uporabnik. Vsa števila naj program tudi izpiše*

Rešitev 9 *Število, do katerega smo že prešteli, si bomo morali nekam zabeležiti, npr. v spremenljivko z imenom i. Šteti bomo glede na navodila začeli s številom 0. Torej bomo spremenljivko i na začetku postavili na vrednost 0. Končali bomo s številom, ki ga je vnesel uporabnik, recimo n. Pogoji za štetje naprej bo torej i <= n. Znotraj zanke while bomo trenutno število (i) izpisali, poleg tega moramo trenutno število tudi povečati, saj bo sicer pogoj za vedno izpolnjen.*

```

1 n = int(input("Vpiši število do katerega želiš šteti: "))
2 i = 0 # števec s katerim bomo šteli
3 while i <= n: # smo ze prešteli do konca?
4     print(i)
5     i = i + 1 # povečanje števca za 1

```

4.4 Stavek +=

Znotraj zanke smo števec povečali za 1 z izvedbo prireditvenega stavka

```
i = i + 1 # povečanje števca za 1
```

Ker je tak način povečevanja vrednosti zelo pogost, v jeziku Python obstaja bližnjica

```
i += 1 # povečanje števca za 1
```

Bistvo zgornjega stavka je, da izvedemo aritmetično operacijo seštevanja in rezultat priredimo spremenljivki, nad katero smo operacijo izvedli. Na podoben način lahko operator prirejanja = kombiniramo z drugimi aritmetičnimi operatorji in števili:

```
>>> x = 10
>>> x += 1 # povečaj za 1
>>> x
11
>>> x -= 2 # zmanjšaj za 2
>>> x
9
>>> x *= 5 # pomnoži s 5
>>> x
45
>>> x /= 9 # deli z 9
>>> x
5.0
>>> x **= 2 # potenciraj na 2
>>> x
25.0
```

4.5 Mrtva zanka

Kaj pa bi se zgodilo, če bi števec v prejšnjem zgledu znotraj zanke pozabili povečati? Spremenljivka `i` (oziroma števec) bi ostala na vrednosti 0 ne glede na to koliko iteracij zanke bi se izvedlo. To pomeni, da bi bil pogoj za vedno izpolnjen (`True`). Kdaj bi se taka zanka končala? Ker je pogoj vedno resničen, se taka zanka nikoli ne konča in tak program je potrebno končati na silo (v okolju Python je temu namenjena kombinacija tipk `ctrl + c`). Na take stvari moramo torej pri programiranju z zanko `while` paziti. Zanki, ki se nikoli ne konča, pravimo *mrtva zanka* (angl. *deadlock*).

4.6 Stavek `break`

Zanko pa lahko prekinemo tudi drugače kot z neizpolnjenostjo pogoja v glavi zanke. Uporabimo lahko namreč stavek `break`, ki prekine izvajanje zanke brez preverjanja pogoja v glavi zanke. Primer uporabe stavka `break` ponazarja spodnja koda

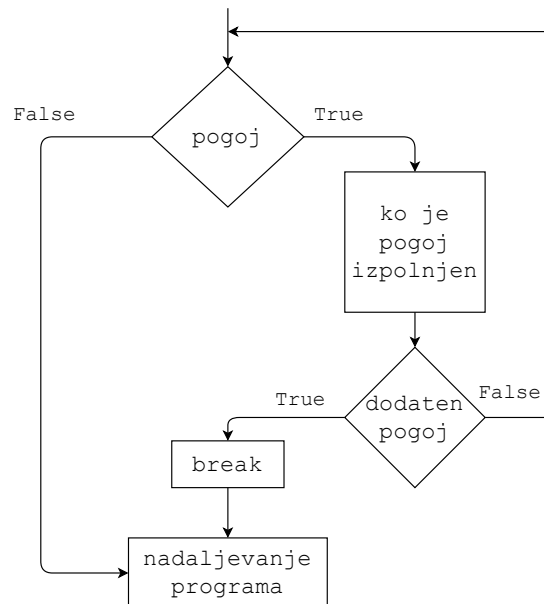
```
while pogoj:
    # telo zanke
    ...
```

```

    if dodaten_pogoj:
        break # prekine izvajanje zanke
# nadaljevanje programa
...

```

Izvedbo primera prikazuje slika 4.3.



Slika 4.3 Primer uporabe stavka `break` znotraj zanke `while`.

Stavek `break` ponavadi uporabljamo v kombinaciji z dodatnim pogojem. V primeru, da je slednji izpolnjen, se izvajanje zanke prekine predčasno. Poglejmo si še en zgled, na katerem bomo potrenirali uporabo stavka `break`.

Zgled 10 Napiši program, ki od uporabnika prebere dve celi števili in izpiše, če sta števili tuji. Števili sta tuji, če nimata nobenega skupnega delitelja, ki je večji od 1.

Rešitev 10 Nalogo bi lahko rešili tako, da bi poiskali največji skupni delitelj podanih števil (naj bosta to števili `st1` in `st2`) in pogledali, če je ta večji od 1. Tokrat se bomo rešitve lotili na malo drugačen način. Preverili bomo, če med kandidati za skupne delitelje obstaja kakšno število, ki deli obe števili, pri čemer bodo kandidati v razponu od števila 2 do manjšega števila, torej do vrednosti `min(st1, st2)`. V primeru, da med kandidati najdemo eno število, ki deli obe podani števili (ostanek po deljenju posameznega števila s kandidatom je enak 0 (`st % delitelj == 0`)), lahko takoj izpišemo, da števili nista tuji.

```

1 st1 = int(input("Vnesi prvo število: "))

```

```

2  st2 = int(input("Vnesi drugo število: "))
3
4  delitelj = 2 # začetna vrednost kandidata
5
6  # kandidat/deljitelj gre do manjšega
7  while delitelj <= min(st1, st2):
8      # ali delitelj deli obe števili?
9      if st1 % delitelj == 0 and st2 % delitelj == 0:
10         print("Števili nista tuji")
11         delitelj += 1

```

Rešitev je še nepopolna, saj izpis poda samo v primeru, ko števili nista tuji. Kako bi lahko program dopolnili, tako da bi izpis podal tudi v primeru, ko sta števili tuji. Tak izpis lahko podamo samo v primeru, ko smo pregledali vse kandidate in nismo našli nobenega, ki deli obe števili. Pomagamo si lahko s pomožno spremenljivko tipa `bool`, v katero bomo shranili informacijo o tem, ali smo že našli kakšnega delitelja. Pri tem bomo na začetku predpostavljali, da delitelja ni. Če ga bomo našli, bomo predpostavko popravili. Na koncu bomo preverili, če smo kakšnega delitelja našli. Če bo odgovor ne (`nasli == False`), bomo izpisali, da sta si števili tuji.

```

1  st1 = int(input("Vnesi prvo število: "))
2  st2 = int(input("Vnesi drugo število: "))
3
4  delitelj = 2 # začetna vrednost kandidata
5  # predpostavljamo, da skupnega delitelja še nismo našli:
6  nasli = False
7
8  # kandidat/deljitelj gre do manjšega
9  while delitelj <= min(st1, st2):
10     # ali delitelj deli obe števili?
11     if st1 % delitelj == 0 and st2 % delitelj == 0:
12         print("Števili nista tuji")
13         nasli = True # popravimo predpostavko
14         delitelj += 1
15
16 # če do tu skupnega delitelja nismo našli, potem ga ni
17 if nasli == False:
18     print("Števili sta tuji")

```

Program sicer deluje pravilno, je pa njegov izpis moteč, v primeru, da najdemo več skupnih deliteljev dveh števil. Vsakič, ko najdemo skupnega delitelja, namreč izpišemo, da smo ga našli. Poleg tega bi lahko izvajanje zanke `while` prekinili takoj, ko smo našli enega skupnega delitelja, saj je to zadosten pogoj, da si števili nista tuji. Uporabimo lahko torej stavek `break`. Končna rešitev bo sledeča.

```

1  st1 = int(input("Vnesi prvo število: "))
2  st2 = int(input("Vnesi drugo število: "))
3
4  delitelj = 2 # začetna vrednost kandidata
5  # predpostavljamo, da skupnega delitelja nismo našli:
6  nasli = False
7
8  # kandidat za skupni delitelj gre do manjšega
9  while delitelj <= min(st1, st2):
10     # ali delitelj deli obe števili?
11     if st1 % delitelj == 0 and st2 % delitelj == 0:
12         print("Števili nista tuji")
13         nasli = True # popravimo predpostavko
14         break # lahko prenehamo z iskanjem
15     delitelj += 1
16
17 # če do tu skupnega delitelja nismo našli, potem ga ni
18 if nasli == False:
19     print("Števili sta tuji")

```

4.7 Veja else

Ena izmed posebnosti jezika Python je tudi to, da lahko vejo `else` uporabljamo tudi v kombinaciji z zanko `while`. Takole:

```

while pogoj:
    # telo zanke
    ...
else:
    # ko pogoj ni več izpolnjen
    ...
# nadaljevanje programa
...

```

Veja `else` se torej izvede, ko pogoj ni več izpolnjen. Vprašanje pa je ali se veja `else` izvede vsakič, ko se izvajanje zanke zaključí? Kakšna je razlika med stavki, ki sledijo zanki `while`, in stavki znotraj veje `else` zanke `while`?

Do bistvene razlike med vejo `else` in običajnimi stavki, ki sledijo zanki `while`, pride, kadar zanko prekinemo s stavkom `break`. V tem primeru namreč skočimo iz zanke, s čimer preskočimo tudi njeno `else` vejo. Slednja se izvede samo v primeru, ko smo zanko prekinili po *običajni* poti, tj. z neizpolnjenostjo pogoja v glavi zanke.

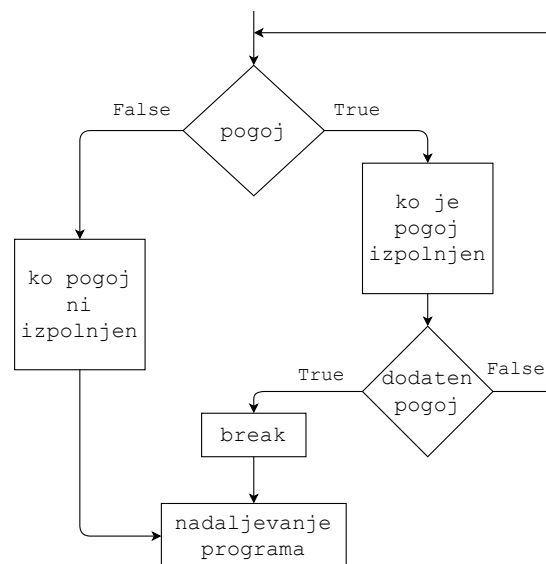
```
while pogoj:
```

```

# telo zanke
...
if dodaten_pogoj:
    break # prekini izvajanje zanke
else: # če zanka ni bila prekinjena z break
    # ko pogoj ni več izpolnjen
    ...
# nadaljevanje programa
...

```

Delovanje zgornjega programa ponazarja slika 4.4.



Slika 4.4 Primer uporabe stavka **break** v kombinaciji z vejo **else** za zanko **while**.

Z vejo **else** lahko določene stavke po zaključku zanke **while** izvedemo samo v primeru, ko zanka ni bila prekinjena s stavkom **break**. Povadimo na prejšnjem zgledu še tole.

Zgled 11 *Napiši program, ki od uporabnika prebere dve celi števili in izpiše, če sta števili tuji. Števili sta tuji, če nimata nobenega skupnega delitelja, ki je večji od 1.*

Rešitev 11 *Del programa, kjer izpisujemo, da si števili nista tuji, bo ostal bolj ali manj nespremenjen. Skrajšamo pa lahko tiste dela programa, ki jih potrebujemo za izpis, da sta si števili tuji. Števili sta si tuji, ko nismo našli nobenega skupnega delitelja. To se zgodi takrat, ko se je zanka **while** odvirtela do konca in nismo našli nobenega skupnega delitelja, torej je posledično nismo prekinili s stavkom **break**. Če zanko **while** dopolnimo z vejo **else**, se bo ta izvedla natanko takrat, ko zanka*

ni bila prekinjena s stavkom **break**, torej takrat, ko nismo našli nobenega skupnega delitelja. Znotraj veje **else** lahko torej zapišemo, da sta si števili tuji. Na tak način se lahko znebimo spremenljivke **nasli** in naredimo program krajši in bistveno bolj pregleden.

```
1 st1 = int(input("Vnesi prvo število: "))
2 st2 = int(input("Vnesi drugo število: "))
3
4 delitelj = 2 # začetna vrednost kandidata
5
6 # kandidat/deljitelj gre do manjšega
7 while delitelj <= min(st1, st2):
8     # ali delitelj deli obe števili?
9     if st1 % delitelj == 0 and st2 % delitelj == 0:
10         print("Števili nista tuji")
11         break # lahko prenehamo z iskanjem
12         delitelj += 1
13 else: # ali se je zanka odvrtela do konca
14     # zanke nismo prekinili s stavkom break
15     print("Števili sta tuji")
```

Osnutek

5 Sezname in metode

5.1 Sekvenčni podatkovni tipi

Podatkovni tipi, ki smo jih srečali do sedaj, so bili večinoma namenjeni temu, da vanje shranimo posamezen (en) podatek. V spremenljivko, ki je pripadala podatkovnemu tipu `int`, smo npr. lahko shranili eno število. V določenih primerih pa se srečamo z veliko količino med seboj podobnih podatkov, nad katerimi želimo izvajati enake ali podobne operacije. V praksi bi to lahko pomenilo, da izvajamo ponavljajoče meritve enake količine, npr. dolžine skoka smučarjev skakalcev. Kaj narediti v takem primeru? Na podlagi našega dosedanjega znanja bi lahko za vsakega skakalca definirali svojo spremenljivko, kar pa ne bi bila ravno najboljša rešitev. Prvi problem tega pristopa bi bil, da je lahko skakalcev zelo veliko. V primeru skakalcev bi bila stvar mogoče še lahko obvladljiva, kaj pa če npr. merimo prisotnost transkriptov genov v celici, ki ima par tisoč genov? Drugi problem je ta, da moramo vsako izmed spremenljivk obravnavati ločeno, kar nas bo pripeljalo do ogromne količine nepregledne *copy-paste* kode. Tretji problem tega pristopa je, da včasih ne vemo čisto točno koliko meritev bomo imeli in koliko spremenljivk bo potrebno definirati (koliko bo skakalcev, koliko je genov v opazovani celici) in zato težko povemo koliko spremenljivk moramo posebej obravnavati. K sreči pa obstajajo t.i. *sekvenčni podatkovni tipi*, v katere lahko shranjujemo večjo količino podatkov oziroma več kot en podatek. Prednost uporabe sekvenčnih podatkovnih tipov je ta, da lahko podatke sproti dodajamo in ne potrebujemo vnaprej definirati števila podatkov, ki jih bomo na koncu imeli. Mimogrede, tudi nizi so sekvenčni podatkovni tipi, saj lahko vanje shranjujemo večjo količino podatkov, ki v tem primeru predstavljajo znake oziroma enočrkovne nize.

5.2 Kaj so sezname?

Drug predstavnik sekvenčnih podatkovnih tipov je seznam oziroma `list`. Za razliko od nizov lahko vanj shranimo poljubne podatke, kot so npr. števila, nizi in tudi drugi sezname. Dodatna prednost uporabe seznamov je ta, da lahko elemente v seznamu dodajamo sproti, zato dolžine seznama ni treba vnaprej definirati. Lahko torej začnemo s praznim seznamom in vsakič, ko dobimo podatek o novi meritvi,

tega v seznam dodamo.

Seznane definiramo z oglatimi oklepaji [in], znotraj katerih naštejemo elemente. Prazen seznam bi naredili takole

```
>>> prazen_seznam = []
```

Seznam, ki vsebuje približno naključne dolžine skokov smučarjev skakalcev pa takole

```
>>> dolzine = [121.4, 143.1, 105.2, 99.3]
```

V isti seznam bi lahko zapisali tudi različne podatkovne tipe, npr. 3 cela števila, 1 decimalko, 5 nizov itd., čeprav v praksi tega ne srečamo pogosto. Ponavadi v seznane shranjujemo podatke, ki pripadajo istemu podatkovnemu tipu, saj se ti podatke nanašajo na ponavljajoče izvajanje npr. določene meritve. Na koncu lahko zato z uporabo seznamov izvedemo določene statistike, npr. kdo je skočil najdlje, kakšna je povprečna dolžina skoka, koliko ljudi je skočilo itd.

5.3 Indeksiranje seznamov

Seznami so urejeni. To pomeni, da je vrstni red, v katerem naštejemo elemente seznama, pomemben. Vsak element v seznamu ima namreč svoj *indeks*. Pri tem se indeksiranje začne z najmanjšim pozitivnim številom, ki v računalništvu ni 1, ampak 0. Indeksi bodo torej šli od števila 0 do dolžine seznama – 1. V primeru zgoraj definirane seznama *dolzine* gredo torej indeksi od 0 do 3, saj seznam vsebuje 4 elemente:

indeksi	0	1	2	3	
dolzine	=	[121.4,	143.1,	105.2,	99.3]

Do elementa na določenem indeksu lahko pridemo z indeksiranjem, ki ga izvedemo tako, da za imenom spremenljivke indeks zapišemo v oglatih oklepajih:

```
ime_seznama[indeks]
```

Do dolžine skoka 0-tega skakalca bi torej prišli takole:

```
>>> dolzine[0]
121.4
```

Kaj pa do zadnjega skakalca? Do dolžine seznama lahko pridemo preko vgrajene funkcije *len*:

```
>>> len(dolzine)
4
```

Funkcijo lahko torej uporabimo pri indeksiranju, kadar ne vemo točno, koliko elementov ima seznam. Do zadnjega elementa torej pridemo takole:

```
>>> dolzine[len(dolzine)-1]
99.3
```

Zakaj moramo od dolžine seznama odšteti 1? Ker smo začeli šteti s številom 0, bo zadnji indeks enak dolžini seznama – 1. Kaj pa če vseeno poskusimo indeksirati po indeksu, ki ga v seznamu ni? V tem primeru seveda dobimo napako:

```
>>> dolzine[len(dolzine)]
IndexError: list index out of range
```

Kot smo do zdaj že večkrat videli ima Python veliko lepih lastnosti. Ena izmed njih je tudi ta, da lahko uporabljamo negativno indeksiranje, pri čemer indeks -1 predstavlja zadnji element, indeks -2 predzadnji in tako naprej. Dolžine skokov imajo torej tudi negativne indekse:

indeksi	-4	-3	-2	-1	
dolzine	=	[121.4,	143.1,	105.2,	99.3]

Prednost takega načina indeksiranja je v tem, da lahko na zelo enostaven način pridemo do zadnjega elementa seznama (brez funkcije `len`):

```
>>> dolzine[-1]
99.3
```

Mimogrede, podobno kot lahko indeksiramo elemente seznamov, lahko indeksiramo tudi elemente nizov. Prav tako lahko dolžino niza preverimo s funkcijo `len`.

```
>>> niz = "banana"
>>> niz[0]
"b"
>>> niz[-1]
"a"
>>> len(niz)
6
```

5.4 Operatorji nad seznamami

Nad seznamami lahko uporabimo različne operatorje, ki smo jih do zdaj uporabljali že npr. nad nizi. Nize smo npr. lahko med seboj seštevali (temu smo sicer rekli konkatencija oziroma lepljenje). Med seboj lahko seštevamo tudi sezname:

```
>>> [1,2,3] + [4,5,6]
[1,2,3,4,5,6]
```

Ne moremo pa seznamom prišteti nečesa, kar ni seznam, npr. števila:

```
>>> [1,2,3]+4
TypeError: can only concatenate list (not "int") to list
```

Lahko pa sezname množimo s celimi števili:

```
>>> [1,2,3]*4
[1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 3]
```

S čim drugim jih ni smiselno množiti, zato Python tega ne podpira:

```
>>> [1,2,3]*[4,5,6]
TypeError: can't multiply sequence by non-int of type 'list'
```

Nad seznamami lahko uporabimo tudi operatorja vsebovanosti `in` in `not in`, ki vrmeta `True` ali `False` v odvisnosti od tega ali je nekaj v seznamu vsebovano ali ne:

```
>>> 1 in [1,2,3]
True
>>> 1 not in [1,2,3]
False
```

Seznane lahko primerjamo z drugimi seznamami z uporabo primerjalnih operatorjev. Takole preverjamo enakost oziroma neenakost dveh seznamov:

```
>>> [1,2,3] == [1,2,3]
True
>>> [1,2,3] != [1,2,3]
False
```

Lahko tudi ugotavljamo, če je prvi seznam manjši od drugega (besedico manjši bi lahko zamenjali tudi z večji, manjši ali enak ter večji ali enak):

```
>>> [1,2,3] < [1,2,3,4]
True
>>> [1,3,3] < [1,2,3]
False
```

Primerjalni operatorji nad seznamami delujejo podobno kot nad nizi, in sicer se gre za leksikografsko primerjanje. Leksikografsko primerjanje je npr. uporabljeno pri sortiranju besed v slovarju. Deluje tako, da med seboj primerjamo istoležne elemente seznama, dokler ne pridemo do neenakosti oziroma do konca enega izmed obeh seznamov. V zgornjem primeru smo prišli do konca prvega seznama. Ker je nekaj kar ne obstaja načeloma manjše kot nekaj kar obstaja, je Python vrnil, da je prvi seznam manjši od drugega. V drugem primeru se je primerjanje ustavilo pri elementih na indeksu 1, saj sta elementa na tem indeksu različna. Ker 3 ni manjše od 2, je Python ugotovil, da prvi seznam ni manjši od drugega in vrnil rezultat `False`.

5.5 Spreminjanje in brisanje elementov seznama

Videli smo že, da lahko do elementov seznama dostopamo preko indeksiranja. Preko indeksiranja pa lahko vrednosti v seznamih tudi spreminjamo. Kako? Tako, da indeksiranje seznama dopolnimo s prireditvenim stavkom:

```
seznam[indeks] = nova_vrednost
```

Tudi brisanje elementov iz seznama lahko izvajamo s pomočjo indeksiranja, le da tokrat pred indeksiranjem uporabimo besedico **del**:

```
del seznam[indeks]
```

5.6 Vgrajene funkcije nad seznamami

Srečali smo že funkcijo `len`, s pomočjo katere lahko ugotovimo kakšna je dolžina seznama. Nad seznamami pogosto uporabljamo še druge vgrajene funkcije, izmed katerih so pogosto uporabljene `min`, `max` in `sum`.

Funkcija `min` vrne najmanjši, funkcija `max` pa največji element v seznamu glede na relacijo `<`. Zdaj lahko končno ugotovimo kakšna je bila dolžina najdaljšega skoka:

```
>>> max(dolzine)
143.1
```

Izračunamo lahko tudi povprečno dožino skoka

```
>>> sum(dolzine)/len(dolzine)
117.25
```

Nad seznamami lahko uporabimo še druge vgrajene funkcije. Nekatere izmed njih bomo srečali kasneje, druge pa boste zagotovo našli, če se bo takšna potreba pokazala.

5.7 Metode

Nad seznamami lahko torej uporabljamo vgrajene funkcije, ki so pač v Pythonu na voljo. Te funkcije lahko sicer uporabimo na poljubnem podatku, ki ni nujno seznam. Obstaja poseben nabor funkcij, ki pa jih lahko uporabljamo samo nad seznamami. Tem funkcijam pravimo *metode seznamov*. V splošnem se izraz *metode* uporablja za posebno družino funkcij, ki pripadajo določenemu *objektu*. Kaj je objekt ne bomo podrobneje razlagali. Lahko pa povemo, da so seznamski objekti (pravzaprav je skoraj vse v Pythonu objekt). Kakorkoli že metode so tiste funkcije, ki pripadajo določenemu objektu. Do posamezne metode seznama lahko pridemo s spodnjim klicem:

```
ime_seznama.ime_metode(argumenti)
```

Klic metode je torej podoben klicu običajne funkcije, le da moramo pred imenom metode podati ime objekta, preko katerega oziroma nad katerim metodo kličemo, imeni pa ločimo s piko (`.`).

Če delamo v okolju IDLE ali v kakšnem še pametnejšem okolju, nam bo to po izpisu imena seznama in pike podalo nabor metod, ki jih imamo na razpolago. Ko v okolju IDLE npr. napišemo

```
>>> dolzine.
```

se po nekaj sekundah pojavijo imena metod, ki jih lahko nad seznamom uporabimo: `append`, `copy`, `clear` itd.

Metode torej razširjajo vgrajene funkcije okolja Python in so vezane na točko določen podatkovni tip, kateremu izbrani objekt pripada. Če bi npr. enako stvar kot zgoraj poskusili z nizom, bi dobili drug seznam metod, ki jih lahko poženemo nad nizom. Metodam kot argument za razliko od vgrajenih funkcij ni potrebno podati seznama (ali pa niza) nad katerim jih želimo izvesti, saj smo seznam (ali pa niz) podali že pred piko – že s samim klicem smo povedali nad čim želimo metodo pognati. Metode vseeno velikokrat vsebujejo določene argumente, ki pač določijo kaj in kako naj metoda nad izbranim objektom naredi.

Prav tako kot obstaja kar veliko vgrajenih funkcij, obstaja tudi veliko metod nad seznamami. Natančneje si bomo v nadaljevanju tega poglavja pogledali tiste, ki jih uporabljamo pogosteje.

5.8 Dodajanje elementov

Dodajanje elementov v seznam je pogosta operacija, zato jo lahko izvedemo na več načinov. Enega smo pravzaprav že srečali, saj lahko za dodajanje elementov v seznam uporabimo kar operator `+`, ki omogoča lepljenje seznamov. Če želimo element seznamu dodati, bomo obstoječemu seznamu prišteli seznam, ki vsebuje ta element. Takole:

```
seznam = seznam + [element]
```

oziroma malo lepše:

```
seznam += [element]
```

Tole dvojje sicer ni popolnoma enako, ampak zaenkrat recimo, da je bolje uporabiti spodnjo različico.

Elemente lahko v sezname dodajamo tudi preko metode `append` in metode `extend`. Obe metodi bosta dodajali na koncu seznama. Razlika je v tem, da v primeru metode `append` dodajamo en element, zato ta metoda kot argument prejme element, ki ga bomo v seznam dodali. Dodajanje bi torej izvedli takole:

```
seznam.append(element)
```

Metoda sicer ne bo ničesar vrnila, bo pa naš seznam spremenila. Primer uporabe je sledeč:

```
>>> seznam = [1,2,3]
>>> seznam.append(4)
>>> seznam
[1,2,3,4]
```


Podobno lahko uporabimo metodo `extend`, ki v seznam dodaja drug seznam. Kot argument moramo torej metodi `extend` podati seznam, ki ga želimo v obstoječ seznam dodati. Takole:

```
seznam.extend(seznam2)
```

Oziroma na prejšnjem zgledu takole:

```
>>> seznam = [1,2,3]
>>> seznam.extend([4])
>>> seznam
[1,2,3,4]
```

Tokrat smo morali element, ki smo ga v seznam dodajali, zapakirati v drug seznam. Povadimo dodajanje elementov v seznam na praktičnem zgledu.

Zgled 12 *Napiši program, ki ga bo lahko uporabil sodnik smučarskih skokov. Program naj sodnika sprašuje po dolžini skoka. V primeru, da sodnik vnese število večje od 0, naj program to število doda v seznam in sodnika vpraša po novi dolžini. Če sodnik vpiše številko 0, naj program izpiše dolžino najdaljšega skoka in povprečno dolžino skoka.*

Rešitev 12 *Sodnikova števila lahko beremo preko funkcije `input`, katere rezultat moramo pretvoriti v podatkovni tip `float`, saj so dolžine decimalna števila. Beremo dokler sodnik ne vnese števila 0, medtem pa dolžine dodajamo v seznam. Na koncu izračunamo povprečno dolžino skoka, poleg tega pa izpišemo tudi najdaljši skok. Program bo sledeč:*

```
1 dolzine = [] # na začetku ni nobene dolžine
2 d = float(input("Vpiši dolžino: ")) # prvo branje
3 while d > 0: # dokler je dolžina veljavna
4     dolzine.append(d) # dodaj dolžino
5     d = float(input("Vpiši dolžino: ")) # ponovno branje
6 print("Najdaljši skok:", max(dolzine))
7 print("Povprečna dolžina:", sum(dolzine)/len(dolzine))
```

Prednost zgornjega programa je v tem, da deluje ne glede na to koliko skokov je v seznamu. Vse dokler sodnik ne vnese kakšne neumnosti.

5.9 Branje seznamov iz ukazne vrstice

Včasih pa bi si želeli celoten seznam prebrati z enim samim uporabnikovim vnosom. Torej bomo spet uporabili funkcijo `input`. Spomnimo se, da funkcija `input` uporabnikov vnos vedno zapiše v niz oziroma v podatkovni tip `str`, ne glede na to kaj je uporabnik vnesel. Tak niz smo v prejšnjih primerih s funkcijo `int`

pretvorili v celo število ali pa s funkcijo `float` v decimalno, če smo želeli podan vnos v nadaljevanju obravnavati kot število. Kaj pa če bi želeli niz, ki ga je vnesel uporabnik, pretvoriti v seznam? Na prvo žogo bi lahko rekli, da pač uporabimo funkcijo `list`. Poskusimo:

```
>>> seznam = list(input("Vnesi seznam: "))
Vnesi seznam: [1,2,3]
>>> seznam
['[', '1', ',', '2', ',', '3', '']
```

To ni ravno tisto, kar smo želeli. Dobili smo namreč seznam vseh znakov, ki v podanem nizu nastopajo (vključno z vejicami in oklepaji).

Kaj bi pravzaprav radi dosegli? To, da se niz, ki ga uporabnik poda funkciji `input` obravnava na enak način, kot če bi isti niz uporabnik vpisal v ukazno vrstico. Temu je namenjena funkcija `eval`, ki kot argument sprejme niz in ga izvede kot kodo v jeziku Python. Poskusimo še to:

```
>>> seznam = eval(input("Vnesi seznam: "))
Vnesi seznam: [1,2,3]
>>> seznam
[1,2,3]
```

V tem primeru stvar deluje, kot bi si želeli. Povadimo še na zgledu.

Zgled 13 *Napiši program, ki ga bo lahko uporabil sodnik smučarskih skokov. Programu naj sodnik poda seznam dolžin smučarskih skokov, program pa naj izpiše dolžino najdaljšega skoka in povprečno dolžino skoka.*

Rešitev 13 *Rešitev bo podobna kot prej, le da tokrat ne potrebujemo zanke.*

```
1 dolzine = eval(input("Vnesi dolžine: "))
2 print("Najdaljši skok:", max(dolzine))
3 print("Povprečna dolžina:", sum(dolzine)/len(dolzine))
```

Slabost programa je ta, da mora sodnik zdaj vse dolžine vnesti naenkrat.

Uporaba funkcije `eval` je sicer lahko v določenih primerih nevarna (če imamo zlobne uporabnike), saj bo slepo izvedla kodo, ki jo bo uporabnik preko vnosa podal.

5.10 Sortiranje seznamov

Zaenkrat znamo določiti najdaljši skok, ne znamo pa določiti najdaljših treh skokov. Najdaljše tri skoke bi lahko poiskali tako, da seznam uredimo (posortiramo), tako da recimo na manjših indeksih vsebuje daljše skoke oziroma, da so skoki urejeni po

dolžini od najdaljšega do najkrajšega. Če razpolagamo s tako urejenim seznamom, lahko za zmagovalne skoke izpišemo skoke na indeksih 0, 1 in 2.

Sortiranje seznamov lahko izvedemo z metodo `sort`:

```
>>> dolzine.sort()
>>> dolzine
[99.3, 105.2, 121.4, 143.1]
```

Metoda `sort` torej sortira seznam, nad katerim smo jo poklicali, in ničesar ne vrača. Opazimo tudi, da je seznam sortirala od najmanjše vrednosti do največje. Najboljše skoke bi torej lahko izpisali tako, da bi izpisali zadnje tri dolžine iz seznama. Lahko pa seznam sortiramo v obratnem vrstnem redu, tako da metodi `sort` nastavimo *opcijski (izbirni)* argument `reverse` na vrednost `True`. Do dokumentacije metode `sort` lahko pridemo preko funkcije `help`:

```
>>> help(list.sort)
Help on method_descriptor:

sort(self, /, *, key=None, reverse=False)
    Stable sort *IN PLACE*.
```

Dokumentacija ni nič kaj preveč izčrpna, vidimo pa lahko, da metoda `sort` sprejema tudi dva opcijska argumenta, in sicer `key`, ki je privzeto enak vrednosti `None` in `reverse`, ki je privzeto enak vrednosti `False`. Opcijski argumenti so tisti argumenti, ki imajo nastavljene *privzete* (angl. *default*) vrednosti. Če ob klicu ne specificiramo drugačnih vrednosti, bodo pač uporabljene privzete vrednosti. Privzete vrednosti pa lahko *povozimo*, tako da specificiramo drugačne vrednosti. Vrstni red urejanja bi lahko spremenili tako, da bi argument `reverse` nastavili na vrednost `True`. V našem primeru takole:

```
>>> dolzine.sort(reverse=True)
>>> dolzine
[143.1, 121.4, 105.2, 99.3]]
```

Rešimo zdaj celoten zgled od začetka do konca.

Zgled 14 *Napiši program, ki ga bo lahko uporabil sodnik smučarskih skokov. Programu naj sodnik poda seznam dolžin smučarskih skokov, program pa naj izpiše najdaljše tri skoke.*

Rešitev 14 *Zdaj bo branju seznama sledilo sortiranje in izpis zmagovalnih dolžin.*

```
1 dolzine = eval(input("Vnesi dolžine: "))
2 dolzine.sort(reverse=True)
3 print("1. mesto", dolzine[0])
4 print("2. mesto", dolzine[1])
5 print("3. mesto", dolzine[2])
```

V dokumentaciji metode `sort` smo videli, da ta sprejema tudi izbirni argument `key`, s katerim lahko določimo, preko katerega naj funkcija sortira. Če bi želeli npr. sortirati seznam po absolutnih vrednostih, bi argumentu `key` priredili funkcijo `abs`. Takole:

```
>>> seznam = [-100, 10, -1, -5, 50]
>>> seznam.sort(key=abs)
>>> seznam
[-1, -5, 10, 50, -100]
```

Seznami (in še kaj drugega) pa bi lahko sortirali tudi preko vgrajene funkcije `sorted`. Ta funkcija deluje na podoben način kot metoda `sort`, le da podanega seznama ne sortira, ampak vrne sortiran seznam. Poglejmo si na zgledu:

```
>>> seznam = [-100, 10, -1, -5, 50]
>>> # kot argument podamo tisto kar želimo posortirati
>>> sorted(seznam)
[-1, -5, 10, 50, -100]
>>> seznam # podan seznam je ostal nespremenjen
[-100, 10, -1, -5, 50]
```

Funkcija torej vrne sortiran seznam, izhodiščni seznam pa je ostal nespremenjen. Kako bi dosegli, da se ime spremenljivke, preko katerega smo funkcijo poklicali, spremeni, tako da vsebuje sortiran seznam? Tako, da bi rezultat funkcije `sorted` priredili spremenljivki:

```
>>> seznam = [-100, 10, -1, -5, 50]
>>> seznam = sorted(seznam)
>>> seznam
[-1, -5, 10, 50, -100]
```

5.11 Seznami seznamov

Vemo že veliko več kot prej, še vedno pa ne vemo kdo je skočil največ in komu moramo podeliti medaljo. Poleg dolžin bi si namreč v ta namen morali beležiti tudi imena tekmovalcev skakalcev. To lahko rešimo tako, da imamo pač dva seznama, tj. seznam dolžin in seznam tekmovalcev. Na istoležnem indeksu imamo v obeh seznamih podatke o istem skakalcu. Takole:

```
>>> dolzine = [121.4, 143.1, 105.2, 99.3]
>>> skakalci = ["Andrej", "Bojan", "Cene", "Dejan"]
```

Andrej je torej skočil 121.4 metra, Dejan pa zgolj 99.3 metra. Zmagovalne tri skoke še vedno lahko dobimo tako, da sortiramo seznam dolžin:

```
>>> dolzine.sort(reverse=True)
>>> dolzine
[143.1, 121.4, 105.2, 99.3]
```

Do problema pa pride, ker zdaj ne vemo več kateremu imenu pripada posamezna dolžina, saj smo indekse v seznamu dolžin s sortiranjem premešali. Kaj lahko naredimo?

Alternativen pristop bi bil, da za beleženje podatkov o dolžinah in imenih uporabimo nov, ugnezden seznam. Torej naredimo seznam seznamov. Takole:

```
>>> skoki = [[121.4, "Andrej"],
              [143.1, "Bojan"],
              [105.2, "Cene"],
              [99.3, "Dejan"]]
```

Kasneje bomo za take primere sicer uporabljali malo drugačno strukturo (zapis podatkov), ampak zaenkrat te še ne poznamo. Naredili smo torej seznam seznamov. Kaj se nahaja v tem primeru na indeksu 0?

```
>>> skoki[0]
[121.4, "Andrej"]
```

Seznam, ki vsebuje podatke o ničtem skakalcu. Kako pa bi prišlo do njegovega imena? Z uporabo verižnega indeksiranja oziroma tako, da po indeksiranju zunanega seznama pač še enkrat indeksiramo notranji seznam:

```
>>> skoki[0][1]
"Andrej"
```

Kaj se bo zgodilo, če tak seznam sortiramo? Nad ugnezdenimi seznamami bo za sortiranje uporabljena relacija $<$, ki smo jo v tem poglavju v povezavi s primerjanjem seznamov že srečali. Rekli smo, da relacija manjše sezname med seboj primerja leksikografsko. Najprej primerja ničti element prvega seznama z ničtim drugega. Če sta enaka, primerja prvi element prvega seznama s prvim elementom drugega seznama in tako naprej. Če bomo torej v ugnezdene sezname na ničto mesto dali dolžine na prvo mesto pa imena, bo sortiranje izvedeno po dolžinah. Po imenih bo sortiranje potekalo samo v primeru, če bosta dolžini pri dveh podseznamih enaki. Poskusimo:

```
>>> skoki.sort(reverse=True)
>>> skoki
[[143.1, 'Bojan'], [121.4, 'Andrej'], [105.2, 'Cene'],
 [99.3, 'Dejan']]
```

Ker smo zdaj sortirali dolžine skokov skupaj z imeni tekmovalcev, informacije o tem kdo je skočil koliko nismo izgubili in lahko izpišemo zmagovalce, ki se nahajajo v prvih treh podseznamih na indeksu 1:

```
>>> skoki [0,1]
'Bojan'
>>> skoki [1,1]
'Andrej'
>>> skoki [2,1]
'Cene'
```

To so zmagovalci. Zapišimo celoten zgled.

Zgled 15 *Napiši program, ki ga bo lahko uporabil sodnik smučarskih skokov. Program naj sodnika sprašuje po dolžini skoka in imenu tekmovalca. V primeru, da sodnik za dolžino vnese število večje od 0, naj program dolžino in ime doda v seznam. Če sodnik vpiše številko 0, naj program izpiše zmagovalce in dolžine njihovih skokov.*

Rešitev 15 *Ponovno bomo brali dolžino po dolžino, le da bomo tokrat v primeru, ko bo vnesena dolžina večja kot 0, prebrali še ime tekmovalca. Potem bomo oboje dodali v seznam skokov. Pomembno je, da na ničto mesto v podseznamu shranimo dolžino skoka, saj želimo podseznime sortirati po dolžini skokov. Na koncu skoke sortiramo in izpišemo zmagovalce in dolžine zmagovalnih skokov.*

```
1 skoki = [] # na začetku ni nobenega skoka
2 d = float(input("Vpiši dolžino: ")) # prvo branje
3 while d > 0: # dokler je dolžina veljavna
4     # branje imena
5     ime = input("Vpiši ime tekmovalca: ")
6     # dodajanje podseznama
7     skoki.append([d,ime])
8     # ponovno branje
9     d = float(input("Vpisi dolžino: "))
10 skoki.sort(reverse=True) # sortiranje
11 print("1. mesto:", skoki[0][1],
12       ", dolžina skoka:", skoki[0][0])
13 print("2. mesto:", skoki[1][1],
14       ", dolžina skoka:", skoki[1][0])
15 print("3. mesto:", skoki[2][1],
16       ", dolžina skoka:", skoki[2][0])
```

5.12 Generiranje seznamov s funkcijo range

Do zdaj smo sezname generirali oziroma dopolnjevali na podlagi vrednosti, ki jih je podal uporabnik. Taki sezname so lahko vsebovali poljubne elemente – pač tisto, kar je uporabnik vnesel.

V določenih primerih želimo imeti sezname s števili v podanem razponu. Praktično uporabo takih seznamov bomo podrobneje spoznali v naslednjem poglavju, zaenkrat pa si pogledjmo, kako jih lahko generiramo.

Generiranje seznamov v podanem razponu omogoča vgrajena funkcija `range`. Funkcijo `range` lahko uporabljamo na tri različne načine, in sicer preko podajanja sledečih argumentov:

- **start**: določa začetek seznama (celo število),
- **stop**: določa konec seznama (celo število),
- **step**: določa korak (celo število).

Pri prvem načinu uporabe funkciji `range` podamo zgolj argument `stop`. V tem primeru bomo dobili seznam vrednosti od 0 do argumenta `stop-1` s korakom 1. Poskusimo:

```
>>> razpon = range(10)
>>> razpon
range(0, 10)
```

Tale izpis je malo čuden. Poklicali smo funkcijo `range` in dobili `range`. Pogledjmo si kakšen je podatkovni tip rezultata:

```
>>> type(razpon)
<class 'range'>
```

Rezultat funkcije `range` torej pripada podatkovnemu tipu oziroma *razredu* `range`, ki nam vrednosti iz razpona vrača sproti, ko jih pač potrebujemo. Zakaj tako? Funkcija `range` je varčna in ponavadi ni nobene potrebe po tem, da bi morali celoten razpon ustvariti naenkrat, ampak naenkrat potrebujemo samo en element iz razpona. S tem, ko funkcija `range` ustvari objekt, ki nam po potrebi vrne želeni element, varčuje tako s procesorjevim časom (hitrost), saj je generiranje dolgih seznamov zamudno, kot tudi s pomnilniškim prostorom, saj dolgi sezname zasedejo veliko prostora. Še vedno pa lahko nad rezultatom funkcije `range` delamo podobne stvari, kot nad seznamami. Lahko jih npr. indeksiramo:

```
>>> razpon[2]
2
```

Ne moremo pa nad njimi klicati metod, ki so definirane nad običajnimi seznamami:

```
>>> razpon.sort()
AttributeError: 'range' object has no attribute 'sort'
```

Poleg tega funkcija `print`, kot smo videli že zgoraj, ne izpiše vrednosti elementov v razponu. Če bi želeli preko funkcije `range` dobiti običajen seznam, lahko uporabimo pretvorbo v seznam preko funkcije `list`:

```
>>> razpon = range(10)
>>> seznam = list(razpon)
>>> seznam
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
>>> type(seznam)
<class 'list'>
```

Za to pa ponavadi ni potrebe. Primerjajmo ekonomičnost funkcije `range` in generiranja običajnega seznama. To lahko poskusimo tako, da s funkcijo `range` naredimo nek relativno velik razpon števil. Npr. od 0 do 10^8-1 :

```
>>> razpon = range(10**8)
```

Tudi če imate počasen računalnik, bo generiranje razpona narejeno v trenutku. Zdaj pa poskusimo iz tega razpona narediti klasičen seznam:

```
>>> seznam = list(razpon)
```

Če vam Python ni javil napake `MemoryError`, je tole verjetno nekaj časa trajalo. Če ni in vas nisem prepričal, poskusite stvar ponoviti z večjim številom, npr. 10^{10} . Vrnimo se k osnovni uporabi funkcije `range`. Mogoče se sprašujete zakaj razpon ne vključuje vrednosti `stop`. Razlogov za to je več. Zaenkrat podajmo najbolj očitnega. Ker funkcija `range` začne šteti z vrednostjo 0 (in ne z 1), bo razpon, ki ga bo vračala, vseboval točno `stop` elementov. Če bi funkciji `range` za argument `stop` podali dolžino nekega seznama, bi razpon vseboval vse indekse tega seznama (ena izmed možnih uporab funkcije `range` se že počasi odkriva).

S funkcijo `range` lahko generiramo razpon elementov, ki se ne začne s številom 0. V tem primeru bomo funkciji poleg argumenta `stop` podali še argument `start`. Najprej seveda navedemo `start`, potem pa `stop`:

```
>>> range(start, stop)
```

Če bi npr. želeli generirati seznam v razponu od 5 do 10, bi napisali takole

```
>>> razpon = range(5, 10)
```

Poglejmo si seznam, ki ga s takim razponom dobimo:

```
>>> list(razpon)
[5, 6, 7, 8, 9]
```

Seznam torej vsebuje argument `start`, argumenta `stop` pa ne. Podobno kot prej. Razpon od 0 do 10 (brez števila 10) bi torej lahko dobili tudi takole:

```
>>> razpon = range(0, 10)
>>> list(razpon)
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```


Do zdaj je bil korak med sosednjima elementoma v razponu vedno enak. To lahko spremenimo tako, da podamo še argument `step`. V tem primeru bomo funkcijo poklicali takole:

```
>>> range(start, stop, step)
```

Argument `step` je opcijski, njegova privzeta vrednost pa je 1. Lahko ga nastavimo na kaj drugega, npr. na 2. Če bi hoteli zgenerirati seznam lihih števil v razponu od 0 do 100, bi to lahko naredili takole:

```
>>> razpon = range(1, 101, 2)
>>> list(razpon)
[1, 3, 5, ..., 97, 99]
```

Zakaj smo argument `start` postavili na 1? Če bi začeli šteti z 0, bi dobili seznam sodih števil.

```
>>> razpon = range(0, 101, 2)
>>> list(razpon)
[0, 2, 4, ..., 98, 100]
```

Korak lahko nastavimo tudi na negativno vrednost:

```
>>> razpon = range(0, 101, -2)
>>> list(razpon)
[]
```

Tokrat smo dobili prazen seznam. Zakaj? Negativen korak pomeni, da štejemo navzdol. Torej mora imeti argument `start` večjo vrednost kot argument `stop`:

```
>>> razpon = range(101, 0, -2)
>>> list(razpon)
[101, 99, 97, ..., 3, 1]
```

Spet smo dobili seznam lihih števil. Zakaj? Zato ker smo začeli šteti z lihim številom. Poleg tega razpon zdaj vključuje število 101, ker je argument `start` v razponu vključen. Razpon sodih števil bi dobili takole

```
>>> razpon = range(100, 0, -2)
>>> list(razpon)
[100, 98, 96, ..., 4, 2]
```

Število 0 tokrat v razponu ni vključeno, ker razpon argumenta `stop` ne vključuje.

5.13 Rezine

V določenih primerih želimo namesto indeksiranja enega samega elementa izvesti indeksiranje razpona elementov v seznamu. Dobimo torej kos oziroma *rezino* (angl.

slice) seznama. Razpon seznama podamo na zelo podoben način, kot smo ga uporabljali pri funkciji `range`, in sicer preko začetka (`start`) rezine, konca (`stop`) rezine in koraka *rezinjenja* (`step`).

Podamo lahko samo začetek rezine: `seznam[start:]`. V tem primeru bo rezina odrezana do konca seznama. Če bi npr. radi dobili vse elemente seznama od vključno petega indeksa naprej, bi napisali takole:

```
>>> seznam = list(range(10))
>>> seznam[5:]
[5, 6, 7, 8, 9]
```

Izhodiščni seznam je kot pri običajnem indeksiranju ostal nespremenjen:

```
>>> seznam
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```

Podamo lahko samo konec rezine: `seznam[:stop]`. V tem primeru se bo začela na začetku seznama in zaključila na indeksu `stop - 1`. Podobno kot pri funkciji `range` tudi pri rezinah `stop` ni vključen v razpon. Če bi npr. radi dobili vse elemente seznama od začetka do petega indeksa (pri tem peti indeks ne bo vključen), bi napisali takole:

```
>>> seznam = list(range(10))
>>> seznam[:5]
[0, 1, 2, 3, 4]
```

Z ne vključenostjo indeksa `stop` smo zopet prišli do točno `stop` vrednosti, saj se štetje začne z indeksom 0.

Pri rezinjenju lahko podajamo tudi zgolj korak: `seznam[::step]`. V tem primeru bo rezina odrezana od začetka do konca seznama, pri čemer bo uporabljen podan korak. Če bi npr. hoteli dobiti vsak drugi element seznama, bi napisali takole:

```
>>> seznam = list(range(10))
>>> seznam[::2]
[0, 2, 4, 6, 8]
```

Korak je lahko tudi negativen. Če bi kot korak npr. napisali vrednost `-1`, bi s tem seznam obrnili. S tem smo namreč povedali, da gremo čez cel seznam s korakom `-1`, torej od konca do začetka:

```
>>> seznam = list(range(10))
>>> seznam[::-1]
[9, 8, 7, 6, 5, 4, 3, 2, 1, 0]
```

Vse zgoraj naštetje kombinacije lahko seveda po mili volji kombiniramo. Če bi npr. hoteli vzeti vsak tretji element seznama v razponu od 2 do 9, bi napisali takole:

```
>>> seznam = list(range(10))
```

```
>>> seznam[2:9:3]
[2, 5, 8]
```

Če je korak negativen, moramo zopet paziti na to, da ima začetek (**start**) večjo vrednost od konca (**stop**)

```
>>> seznam = list(range(10))
>>> seznam[2:9:-1]
[]
>>> seznam[9:2:-1]
[9, 8, 7, 6, 5, 4, 3]
```

Če bi želeli iti od konca, do nekega indeksa proti začetku, bi to lahko podali kot `seznam[:stop:-1]`, npr.

```
>>> seznam = list(range(10))
>>> seznam[:1:-1]
[9, 8, 7, 6, 5, 4, 3, 2]
```

stop tudi tokrat ni vključen.
Povadimo rezine še na enem zgledu.

Zgled 16 *Napiši program, ki ga bo lahko uporabil sodnik smučarskih skokov. Programu naj sodnik poda seznam dolžin smučarskih skokov, program pa naj izpiše najdaljše tri skoke.*

Rešitev 16 *Podobno kot prej bomo seznam prebrali in uredili. Zmagovalce lahko zdaj izpišemo v eni vrstici. Tokrat bo program za razliko od prej deloval tudi v primeru, če bo sodnik vnesel manj kot 3 skoke. Rezine pač režejo dokler gre in primeru da razpon preseže indekse seznama, napake ne javljajo.*

```
1 dolzine = eval(input("Vnesi dolžine: "))
2 dolzine.sort(reverse=True)
3 print(dolzine[:3])
```

5.14 Indeksiranje nizov

Kot smo že omenili lahko podobno kot sezname indeksiramo tudi nize. Prav tako lahko nad nizi izvajamo rezine. Povadimo najprej rezinjenje.

Zgled 17 *Napiši program, ki bo od uporabnika prebral dve zaporedji nukleotidnih baz (zapisani kot niza) in med njima na sredini izvedel križanje, tako da bo sestavil dve novi zaporedji nukleotidnih baz in jih izpisal.*

Rešitev 17 Program bo torej od uporabnika prejel dva niza. Najprej bomo določili indeksa, kjer bomo križanje naredili. To bo na polovici posameznega zaporedja. Dolžino posameznega zaporedja bomo delili z 2, pri čemer bomo uporabili celoštevilsko deljenje (`//`), saj morajo biti indeksi cela števila. Potem bomo odrezali rezine in jih med seboj sestavili (z operatorjem lepljenja `+`) ter izpisali.

```

1  gen1 = input("Vpiši prvo zaporedje: ")
2  gen2 = input("Vpiši drugo zaporedje: ")
3
4  # kje prerežemo gen 1?
5  i1 = len(gen1)//2 # celoštevilsko deljenje z 2
6  # kje prerežemo gen 2?
7  i2 = len(gen2)//2 # celoštevilsko deljenje z 2
8
9  gen11 = gen1[:i1] # prva polovica gena 1
10 gen12 = gen1[i1:] # druga polovica gena 1
11 gen21 = gen2[:i2] # prva polovica gena 2
12 gen22 = gen2[i2:] # druga polovica gena 2
13
14 # lepljenje iz izpis
15 print(gen11 + gen22)
16 print(gen21 + gen12)

```

Iz zgornjega zgleda vidimo še eno prednost tega, da `stop` v rezino ni vključen. Če prvo rezino režemo do indeksa `stop`, drugo pa od istega indeksa naprej, bosta rezini nepresečni, kar pomeni, da element na indeksu `stop` ne bo podvojen. Povadimo zdaj še običajno indeksiranje, ki ga bomo potem pohitrili z rezinami.

Zgled 18 Palindrom je niz, ki se na enak način bere naprej kot nazaj. Napiši program, ki od uporabnika prebere niz in izpiše, če podani niz je oziroma ni palindrom.

Rešitev 18 Prva rešitev bo temeljila na zanki `while`, s katero se bomo sprehajali od začetka proti koncu niza. Zanko `while` lahko torej ponavljamo, dokler z nekim števcem (npr. `i`) ne preštejemo do konca niza. Začeli bomo pa seveda na začetku, torej pri vrednosti 0 (`i=0`). V zanki `while` bomo primerjali enakost znaka na indeksu `i` z znakom na indeksu `-i-1`. Če se bodo ti pari ujemali **povsod**, bomo lahko sklepali, da je niz palindrom. Takoj, ko bomo našli **en** primer, kjer se par ne ujema (protiprimer), pa bomo lahko sklepali, da niz ni palindrom.

```

1  niz = input("Vpiši niz: ")
2  i = 0 # začeli bomo na začetku niza
3  while i < len(niz): # do konca niza
4      if niz[i] != niz[-i-1]: # protiprimer

```

```

5         print("Niz ni palindrom")
6         break
7     i += 1 # gremo na naslednji par
8 else: # če smo prisli do konca brez break-a
9     print("Niz je palindrom")

```

Program bi sicer lahko nekoliko pohitrili, saj se nam ni treba premikati do konca niza, ampak je dovolj, da končamo, ko števec *i* pride do polovice niza. Pogoji v zanki *while* bi torej lahko spremenili v *i < len(niz)//2*.

Do bistveno lepše rešitve pa pridemo, če uporabimo rezine. Niz je palindrom, če se bere naprej enako kot nazaj. Torej mora biti naprej prebran niz (*niz*) enak nazaj prebranemu niz (*niz[::-1]*). Program je torej sledeč:

```

1 niz = input("Vpiši niz: ")
2 if niz == niz[::-1]:
3     print("Niz je palindrom")
4 else:
5     print("Niz ni palindrom")

```

5.15 Sprehajanje čez sezname

Do zdaj smo se temu sicer izogibali, ampak pri delu s seznamami je ena izmed najpogostejših operacij sprehajanje nad seznamami. Kako narediti tak sprehod? Zgoraj smo se z zanko *while* sprehajali nad indeksi niza. Podoben sprehod bi lahko naredili tudi nad seznamami. Posamezen element seznama bi lahko izpisali npr. takole:

```

i = 0
while i < len(seznam):
    print(seznam[i])
    i += 1

```

Sprehajamo se torej po indeksih od začetka (0) do konca seznama (*len(seznam)-1*). Zgornja koda je sicer popolnoma pravilna, ni pa najlepša, saj je sprehajanje nad seznamami in seznamu podobnimi podatki v Pythonu namenjena posebna zanka, zanka *for*.

Osnutek

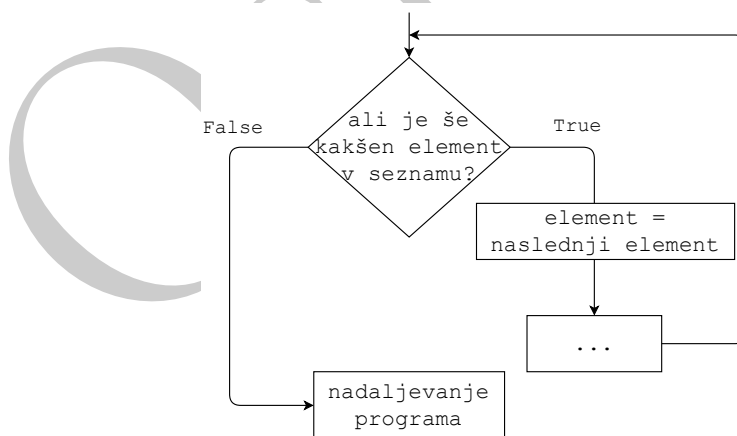
6 Zanka for

6.1 Sprehajanje čez seznane z zanko for

Kot smo videli na koncu prejšnjega poglavja, se lahko čez seznam (ali niz) sprehodimo z uporabo zanke `while`, pri čemer sprehod vršimo preko indeksov seznama (ali niza). Preko indeksov lahko potem posredno pridemo tudi do vrednosti elementov seznama (ali niza). Veliko bolj elegantno pa se čez seznam (ali pač niz) sprehodimo z uporabo zanke `for`:

```
for element in seznam:  
    # telo zanke  
    # spremenljivka element vsebuje trenutni element  
    ...  
# nadaljevanje programa  
...
```

Potek izvedbe osnovne oblike zanke `for` ponazarja slika 6.1 Izvajanje zanke torej



Slika 6.1 Potek izvedbe osnovne oblike zanke `for`.

ponavljamo vse dokler je v seznamu (ali nizu) še kakšen element, pri čemer se spremenljivka `element` pomika od začetka proti koncu seznama (ali niza). Če bi

npr. želeli izpisati vse elemente v seznamu, pri čemer bi vsak element izpisali v svoji vrstici, bi to lahko naredili takole:

```
>>> seznam = [1,2,3]
>>> for element in seznam:
        print(element)

1
2
3
```

Na podoben način bi se lahko sprehodili tudi čez niz:

```
>>> niz = "ABC"
>>> for znak in niz:
        print(znak)

A
B
C
```

V tem primeru se zanka for torej sprehaja čez znake niza. Povadimo sprehajanje še na zgledu.

Zgled 19 *Napiši program, ki od uporabnika preko funkcije `eval` prebere seznam in izpiše najmanjši element seznama (brez uporabe funkcije `min`).*

Rešitev 19 *Najmanjši element bomo našli tako, da se bomo z zanko `for` sprehodili čez seznam in si zapomnili element, ki je pač najmanjši. Kako pa vemo, da je nek element najmanjši, če ostalih še nismo pregledali? Težko. Vemo pa, če je nek element manjši od vseh elementov, ki smo jih pregledali preden smo do njega prišli. Nalogo lahko rešimo tako, da naredimo predpostavko, da je najmanjši ničti element v seznamu. Potem naredimo sprehod čez celoten seznam. Če bomo našli kakšen element, ki je manjši od trenutno najmanjšega, bomo trenutno najmanjši element postavili na tega, ki je očitno manjši. To bomo nadaljevali, dokler ne pridemo do konca seznama.*

```
1  seznam = eval(input("Vnesi seznam: "))
2
3  najmanjsi = seznam[0] # trenutno najmanjši
4
5  for element in seznam: # sprehod čez elemente
6      if element < najmanjsi: # našli manjšega?
7          najmanjsi = element # popravimo vrednost
8
9  print(najmanjsi)
```


Program sicer deluje pravilno, ampak bi ga lahko še nekoliko optimizirali. Trenutno namreč ničti element v seznamu pregledamo dvakrat. Sprehoda z zanko `for` nam torej ne bi bilo potrebno delati čez cel seznam, ampak bi ga lahko naredili čez rezino seznama, ki se začne na indeksu 1. Torej bi zanko `for` lahko delali čez rezino `seznam[1:]`.

6.2 Sprehajanje s funkcijo `range` in sprehajanje čez indekse

Z zanko `for` se lahko sprehajamo tudi čez sezname, ki jih generira funkcija `range`. Na tak način se lahko sprehajamo čez vrednosti elementov v določenem razponu. Vsa števila od 0 do vključno števila, ki ga je vnesel uporabnik, bi torej lahko izpisali tudi takole:

```
n = int(input("Vnesi število: "))
for i in range(n+1):
    print(i)
```

Zanko `for` bi lahko na podoben način uporabili za sprehajanje po indeksih seznama. Indekse in vrednosti v seznamu na posameznih indeksih bi lahko izpisali takole:

```
for i in range(len(seznam)):
    print(i, seznam[i])
```

Tokrat funkciji `range` kot argument `stop` podamo dolžino seznama, kar pomeni, da bo funkcija zgenerirala razpon elementov v intervalu od 0 do `len(seznam)-1`, kar je ravno razpon indeksov seznama. Tudi zato torej argument `stop` v interval ni vključen in zato funkcija `range` (tudi) deluje kakor deluje.

6.3 Sprehajanje čez elemente ali čez indekse?

Zgornji program bo poleg indeksa izpisal še vrednost elementa, ki se nahaja na posameznem indeksu. Ali bi lahko do indeksa elementov prišli tudi v primeru, ko se sprehajamo neposredno po elementih seznama? Težko. Zato v primeru, ko informacijo o indeksu potrebujemo, izvajamo zanko čez indekse in ne čez elemente. Poglejmo si spodnji primer, kjer rešitev zahteva izvedbo sprehoda čez indekse seznama.

Zgled 20 *Napiši program, ki od uporabnika preko funkcije `eval` prebere seznam in izpiše najmanjši element seznama ter njegov indeks.*

Rešitev 20 *Najmanjši element bomo našli na podoben način kot prej, le da si moramo tokrat zapomniti tudi njegov indeks. Ker preko direktnega sprehoda čez elemente seznama informacije o indeksih elementov nimamo, se bomo morali sprehoditi čez indekse seznama.*

```

1 seznam = eval(input("Vnesi seznam: "))
2
3 najmanjsi = seznam[0] # trenutno najmanjši element
4 najmanjsi_i = 0 # zapomnimo si tudi njegov indeks
5
6 for i in range(len(seznam)): # sprehod čez indekse
7     element = seznam[i] # preko indeksa do elementa
8     if element < najmanjsi: # našli manjšega?
9         najmanjsi = element # popravimo vrednost
10        najmanjsi_i = i # popravimo indeks
11
12 print(najmanjsi)
13 print(najmanjsi_i)

```

Spet bi lahko pri sprehodu prvi element seznama izpustili, tako da bi se sprehajali čez razpon indeksov `range(1, len(seznam))`.

Zgornja rešitev ima manjšo pomanjkljivost, in sicer ne upošteva, da se lahko enako majhen element v seznamu pojavi večkrat. V tem primeru vrne zgolj indeks njegove prve pojavitve. Naprednejšo rešitev prikazuje spodnji zgled.

Zgled 21 *Napiši program, ki od uporabnika preko funkcije `eval` prebere seznam in izpiše najmanjši element seznama ter vse indekse njegove pojavitve.*

Rešitev 21 *Rešitev bo podobna kot prej, le da si bomo indekse pojavitve najmanjšega elementa zabeležili kar v seznam. V primeru, da bomo našli manjši element od trenutnega, bomo naredili nov seznam, ki bo vseboval samo en indeks (trenutni indeks). V primeru, da bomo našli element, ki bo enak trenutno najmanjšemu, bomo v seznam indeksov dodali trenutni indeks. V prejšnjih dveh zgledih smo na koncu omenili boljšo rešitev, ki pri sprehodu izpusti ničti element seznama, saj smo tega upoštevali že pred zanko. Tokrat bo program brez te "optimizacije" deloval narobe. V primeru, da bo najmanjši element na ničtem mestu, bo njegov indeks v seznamu najmanjših indeksov namreč nastopal dvakrat.*

```

1 seznam = eval(input("Vnesi seznam: "))
2
3 najmanjsi = seznam[0] # trenutno najmanjši element
4 najmanjsi_i = [0] # v seznam shranimo njegov indeks
5
6 # sprehod čez indekse (ničti element izpustimo)
7 for i in range(1, len(seznam)):
8     element = seznam[i] # preko indeksa do elementa
9     if element < najmanjsi: # našli manjšega?

```

```

10     najmanjsi = element # popravimo vrednost
11     najmanjsi_i = [i] # resetiramo seznam indeksov
12     elif element == najmanjsi: # našli enako majhnega
13         najmanjsi_i.append(i) # dodamo indeks
14
15     print(najmanjsi)
16     print(najmanjsi_i)

```

6.4 Spreminjanje elementov seznama z zanko for

Kaj pa v primeru da želimo seznam v zanki spremeniti, npr. da želimo vse negativne vrednosti seznama spremeniti v pozitivne (izračunati želimo absolutne vrednosti elementov seznama in seznam skladno s tem posodobiti). Žal funkcije `abs` nad seznamom ne moremo pognati (napaka), zato moramo izračunati absolutno vrednost vsakega elementa posebej, pri čemer lahko to rešimo z uporabo zanke `for`. Poskusimo z običajnim sprehodom čez elemente seznama.

```

>>> seznam = [-1, 10, -5, 15, 0, -3]
>>> for element in seznam:
        element = abs(element)
        print(element)
1
10
5
15
0
3
>>> print(seznam)
[-1, 10, -5, 15, 0, -3]

```

Elemente smo torej uspešno postavili na njihove absolutne vrednosti, na kar nakazujejo izpisi, ki smo jih izvedli v telesu zanke. Kot pa vidimo iz izpisa, ki je sledil zanki, se seznam ni spremenil, saj še vedno vsebuje negativne elemente. V konkretnem primeru torej samega seznama nismo spreminjali. Če bi želeli spreminjati seznam, bi to lahko naredili preko indeksiranja:

```

>>> seznam = [-1, 10, -5, 15, 0, -3]
>>> for i in range(len(seznam)):
        seznam[i] = abs(seznam[i])
        print(seznam[i])
1
10
5

```

```

15
0
3
>>> print(seznam)
[1, 10, 5, 15, 0, 3]

```

Zdaj se je seznam seveda spremenil, saj smo absolutne vrednosti direktno prirejali seznamu na posameznem indeksu.

6.5 Zanka `for` ali zanka `while`?

Vidimo, da so naši programi z uporabo zanke `for` v določenih veliko krajši in lepši kot v primeru uporabe zanke `while`. Poleg tega nam pri uporabi zanke `for` ni potrebno skrbeti, da bo program za vedno obtičal v zanki (mrtva zanka). Zakaj bi torej sploh uporabljali zanko `while`? Izkaže se, da je zanka `while` bolj splošna kot zanka `for` in da lahko z njo rešimo določene probleme, ki jih z zanko `for` ne moremo. Kako bi npr. z zanko `for` od sodnika smučarskih skokov brali dolžine skokov, dokler sodnik ne vnese števila 0? Koliko ponovitev bi morali narediti? Kako bi z zanko `for` odštevali manjše število od večjega, dokler števili ne bi postali enaki? Odgovor je enostaven. Težko.

Vprašajmo se, kaj je skupnega primerom, kjer zanka `for` odpove. V obeh zgornjih primerih je število ponovitev, ki jih bo morala zanka narediti, vnaprej težko predvidljivo. V splošnem velja, da zanko `while` uporabljamo, kadar število ponovitev zanke težko podamo vnaprej, lahko pa oblikujemo pogoj, ki bo določil, do kdaj naj se zanka izvaja. V primeru, da je število ponovitev predvidljivo (npr. podan je razpon štetja ali pa seznam s fiksno dolžino, čez katerega se sprehajao) pa je kot nalašč zanka `for`.

6.6 Stavek `break`

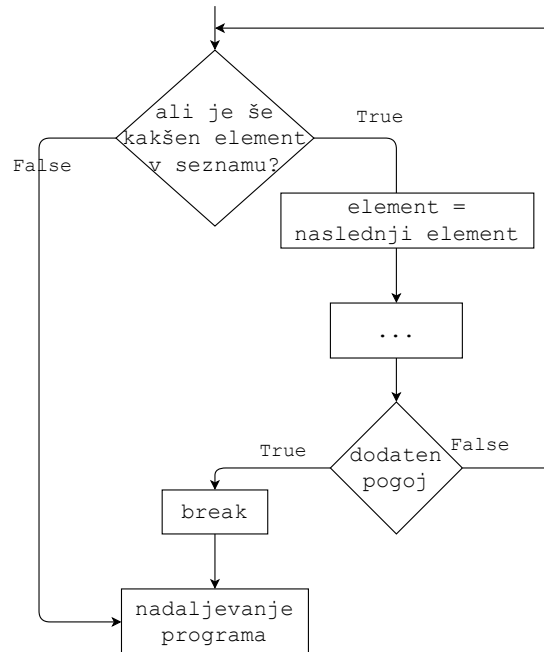
V kombinaciji z zanko `for` lahko prav tako kot pri zanki `while` uporabljamo stavek `break`. Ta izvajanje zanke prekine, kljub temu, da ta še ni prišla do konca seznama (ali česa drugega). Primer uporabe stavka `break` znotraj zanke `for` ponazarja spodnja koda:

```

for element in seznam:
    # telo zanke
    ...
    if dodaten_pogoj:
        break # prekine izvajanje zanke
# nadaljevanje programa
...

```

Potek izvedbe kode iz primera prikazuje slika 6.2.



Slika 6.2 Potek izvedbe zanke `for` v kombinaciji s stavkom `break`.

6.7 Veja else

Podobno kot lahko vejo `else` kombiniramo z zanko `while`, jo lahko kombiniramo tudi z zanko `for`:

```

for element in seznam:
    # telo zanke
    ...
    if dodaten_pogoj:
        break # prekini izvajanje zanke
else: # samo če zanka ni bila prekinjena z break
    # konec seznama
    ...
# nadaljevanje programa
...
  
```

Veja `else` se bo kot pri zanki `while` izvedla samo v primeru, ko zanka ni bila prekinjena s stavkom `break`. Demonstrirajmo uporabo koncepta na zgledu s tujimi števili.

Zgled 22 *Napiši program, ki od uporabnika prebere dve celi števili in izpiše, če sta števili tuji. Števili sta tuji, če nimata nobenega skupnega delitelja, ki je večji od 1.*

Rešitev 22 *Program bo strukturiran zelo podobno kot v primeru zanke `while`, le da bomo tokrat razpon števil, čez katera se sprehaja kandidat, ustvarili z uporabo funkcije `range`.*

```

1  st1 = int(input("Vnesi prvo število: "))
2  st2 = int(input("Vnesi drugo število: "))
3
4  # sprehod od 2 do manjšega od obeh števil
5  # desni del intervala naj bo vključen,
6  # zato prištejemo 1
7  for delitelj in range(2, min(st1, st2)+1):
8      # ali delitelj deli obe števili?
9      if st1 % delitelj == 0 and st2 % delitelj == 0:
10         print("Števili nista tuji")
11         break # lahko prenehamo z iskanjem
12 else: # ali se je zanka odvrtela do konca
13     # zanke nismo prekinili s stavkom break
14     print("Števili sta tuji")

```

6.8 Gnezdenje zank

Podobno kot smo gnezdili stavke `if` lahko gnezidmo tudi zanke. To pomeni, da bomo zanko izvajali znotraj druge zanke. Primer gnezdenja zanke `for` prikazuje spodnji izsek kode:

```

>>> for i in range(5):
      for j in range(5):
          print(i,j)

```

0 0
0 1
0 2
0 3
0 4
1 0
1 1
...
3 3
3 4
4 0

```

4 1
4 2
4 3
4 4

```

Notraja zanka `for` torej za vsako iteracijo zunanje zanke izvede enako število ponovitev.

Potrenirajmo na zgledu.

Zgled 23 *Napiši program, ki od uporabnika prebere celo število in izpiše poštrevanko števil od 1 do vključno podanega števila.*

Rešitev 23 Števila od 1 do podanega števila n bomo najprej množili z 1, potem z 2, potem s 3 in tako naprej, dokler ne pridemo do števila n . To lahko enostavno rešimo z uporabo ugnezdene zanke.

```

n = int(input("Vnesi število: "))

for i in range(1, n+1): # zunanja zanka
    for j in range(1, n+1): # notranja zanka
        print(i*j) # izpis produkta
    print() #nova vrstica

```

V primeru, da uporabnik vpiše število 3, bo izpis sledeč:

```

1
2
3

2
4
6

3
6
9

```

V zgornjih primerih je bila notranja zanka neodvisna od tega, kako daleč se je že odvila zunanja zanka. Ponavadi pa temu ni tako. Primer ugnezdene zanke, pri kateri je razpon notranje zanke odvisen od števila izvedenih iteracij zunanje zanke, prikazuje spodnji izsek kode:

```

>>> for i in range(5):
        for j in range(i,5):
            print(i,j)

```

```

0 0
0 1
0 2
0 3
0 4
1 1
1 2
1 3
1 4
2 2
2 3
2 4
3 3
3 4
4 4

```

V prvi iteraciji zunanje zanke, se torej notranja zanka izvede petkrat, v drugi štirikrat, v tretji trikrat, v četrti dvakrat, v peti pa zgolj enkrat. Kasnejša kot je iteracija zunanje zanke, manjše je število ponovitev ugnезdene zanke.

Povadimo tako gnezdenje še na primeru s tujimi števili.

Zgled 24 *Napiši program, ki od uporabnika prebere celo število in izpiše vsa števila, ki so podanemu številu tuja in so od njega manjša.*

Rešitev 24 *Kandidati, ki jih moramo torej obravnavati, se gibljejo v razponu od števila 1 (ki je vsem številom tuje število) do števila $n-1$, pri čemer je n število, ki ga je vnesel uporabnik. Kako za posameznega kandidata preverimo, če je tuj številu n ? Podobno kot prej – tako da se sprehodimo od števila 2, do manjšega od obeh števil. Če smo našli kakšnega delitelja, si števili očitno nista tuji.*

```

1  n = int(input("Vnesi število: "))
2
3  for kandidat in range(1, n): # razpon cez kandidate
4      # kandidat ne sme imeti nobenega skupnega delitelja
5      # ugnездimo kodo iz prejšnjih zgledov
6      st1 = n
7      st2 = kandidat
8
9      # sprehod od 2 do manjšega od obeh števil
10     # desni del intervala naj bo vključen, zato prištejemo 1
11     for delitelj in range(2, min(st1, st2)+1):
12         # ali delitelj deli obe števili?
13         if st1 % delitelj == 0 and st2 % delitelj == 0:

```



```

14         break # lahko prekinemo ugnezdjeno zanko
15     else: # ali se je ugnezdjena zanka odvirtela do konca
16         # ugnezdene zanke nismo prekinili s stavkom break
17         print(kandidat)

```

Ugnezdjena zanka je v tem primeru odvisna od tega kako daleč je naš program prišel z zunanjo zanko. Mimogrede, na podoben način bi lahko gnezdili tudi zanko `while`.

6.9 Izbirni argumenti funkcij in izbirni argumenti funkcije `print`

Tole sicer ni neposredno povezano z zanko `for`, bo pa služilo kot osnova za dopolnitev zgleda s poštevanko.

Povedali smo že, da funkcije sprejemajo argumente, ki jih ob klicu pač podamo. V določenih primerih pa imajo funkcije tudi t.i. *izbirne* ali *opcijske* argumente, za katere velja da imajo (pred)nastavljeno *privzeto* vrednost. V primeru, da vrednosti teh argumentov eksplicitno ne podamo, bodo ti nastavljeni na njihove privzete vrednosti. V primeru, da vrednosti tem argumentom podamo, bomo s tem *povozili* privzete vrednosti in uporabljene bodo podane (naše) vrednosti.

Poglejmo si dva izbirna argumenta funkcije `print` in primer njune uporabe. Do dokumentacije funkcije `print` lahko pridemo preko funkcije `help`:

```

>>> help(print)
Help on built-in function print in module builtins:

print(...)
    print(value, ..., sep=' ', end='\n', file=sys.stdout,
          flush=False)

    Prints the values to a stream, or to sys.stdout by default.
    Optional keyword arguments:
    file: a file-like object (stream); defaults to the current
          sys.stdout.
    sep:   string inserted between values, default a space.
    end:   string appended after the last value, default a newline.
    flush: whether to forcibly flush the stream.

```

Zaenkrat nas bosta zanimala predvsem argumenta `sep` in `end`. Funkcija `print` deluje tako, da sprejme poljubno število števil, nizov in še česa drugega, to med seboj združi in izpiše na zaslon. Pri tem argument `sep` določa, s čim naj podane številke, nize in še kaj drugega med seboj združi. Privzeto je ta argument postavljen

na vrednost ' ', kar vidimo iz zgleda klica funkcije (`sep=' '`). To pomeni, da bo izpis narejen tako, da bodo med podanimi argumenti za izpis vstavljeni presledki. Povadimo:

```
>>> print(1,2,3) # privzeta vrednost argumenta
1 2 3
>>> print(1,2,3, sep='') # brez presledka
123
>>> print(1,2,3, sep='+++') # poljuben niz kot ločilo
1+++2+++3
```

Izbirni argument `end` podaja niz, ki naj se vstavi na koncu izpisa. Privzeto je argument `end` nastavljen na znak `'\\n'` (`end='\\n'`), ki predstavlja znak za novo vrstico (angl. *line feed*). Tudi tega lahko postavimo na kakšno drugo vrednost. Povadimo nastavljanje opsijskih argumentov na zgledu v kombinaciji z ugnézdeno zanko `for`.

Zgled 25 *Napiši program, ki od uporabnika prebere celo število in izpiše poštévanko števil od 1 do vključno podanega števila. Pri tem naj bo poštévanka s posameznim številom podana v svoji vrstici, števila pa naj bodo ločena s presledki.*

Rešitev 25 *Rešitev bo podobna kot prej, le da se tokrat ne bomo pomikali v novo vrstico po vsakem izpisu. To lahko naredimo tako, da opsijski argument `end` nastavimo na znak ' '.*

```
n = int(input("Vnesi število: "))

for i in range(1, n+1): # zunanja zanka
    for j in range(1, n+1): # notranja zanka
        print(i*j, end = ' ') # izpis produkta brez nove vrstice
    print() #nova vrstica
```

V primeru, da uporabnik vpiše število 3, bo tokrat izpis sledeč:

```
1 2 3
2 4 6
3 6 9
```

7 Uporaba in pisanje funkcij

7.1 Kaj so funkcije in zakaj so uporabne?

Kot že vemo, funkcije predstavljajo del kode, ki jo lahko izvedemo tako, da funkcijo pač pokličemo.

Uporaba funkcij ima veliko prednosti. Govorili smo že o tem, da je glavno vodilo programiranja razdelitev problemov na obvladljive podprobleme. Določanje algoritma, ki ga potem samo še prenesemo v programsko kodo, je podobno določanju recepta, ki ga potem prenesemo v okusno jed. Prav tako, kot se moramo pri kuhanju zavedati sestavin, ki jih imamo na razpolago, se moramo tudi pri programiranju zavedati gradnikov programskega jezika, ki jih lahko pri pisanju algoritma uporabimo.

Funkcije nam omogočajo, da osnovne korake za reševanje programa vgradimo v enostavnejše funkcije, enostavnejše funkcije v kompleksnejše in tako naprej. Podobno, kot če bi pri peki torte lahko uporabili že vnaprej pripravljeno testo, preliv in kar se pač pri torti še uporabi, namesto da moramo torto sestaviti iz enostavnejših (nižjenivojskih) sestavin, kot so jajca, mleko in sladkor. Tako, kot bi lahko tudi pri peki seveda šli v drug ekstrem in se lotili reje kokoši, bi na veliko nižji nivo lahko šli tudi pri programiranju, ampak pustimo to za kdaj drugič. S pisanjem svojih funkcij se lahko torej najprej lotimo enostavnejših korakov, ki predstavljajo del rešitve izbranega problema. Potem lahko vmesne rešitve (velikokrat na enostaven način) združimo v končno rešitev. Če bi npr. želeli najti vsa praštevila v določenem razponu števil, bi lahko najprej napisali funkcijo, ki za podano število preveri, če je praštevilo. Vse kar bi morali narediti potem bi bil zgolj klic te funkcije za vsako število z intervala.

Zgled s praštevili pa nam je posredno razodel še eno veliko prednost uporabe funkcij. Isto kodo, tj. preverjanje ali je neko število praštevilo, bomo poklicali večkrat, vsakič seveda z drugim argumentom, tj. številom, ki je kandidat za praštevilo. Funkcije nam torej omogočajo tudi to, da lahko isti kos kode večkrat pokličemo brez tega, da bi jo vključevali v zanke ali pa kopirali v vse dele programa, kjer jo potrebujemo.

To kodo bi lahko delili tudi z drugimi programerji. Če smo npr. napisali zelo dobro funkcijo za iskanje praštevil in smo nad njo nadvse navdušeni, hkrati pa vemo, da bi bila lahko koristna tudi za druge iskalce praštevil, lahko funkcijo enostavno

zapakiramo v t.i. modul, ki ga objavimo na internetu.

7.2 Kako definiramo funkcijo?

Vsaki funkciji, ki jo želimo v naših programih ponovno uporabiti, moramo dati seveda neko ime, preko katerega jo bomo lahko po potrebi poklicali. Skupaj s seznamom argumentov, ki jih bo naša funkcija sprejela, to podamo v definiciji funkcije. Definicijo funkcije začnemo z rezervirano besedo **def** in končamo z dvopičjem:

```
def ime_funkcije(argument_1, argument_2, ..., argument_n):
```

Definiciji funkcije sledi njena vsebina. Stavke, ki so v funkciji vsebovani tudi tokrat določimo z zamikanjem na začetku vrstice (podobno kot pri pogojnemu stavku in zankah). Ko želimo Pythonu sporočiti, da koda ni več del funkcije, enostavno nehamo zamikati.

Spodnji primer predstavlja definicijo enostavne funkcije, ki sešteje vrednosti dveh spremenljivk (*a* in *b*) v novo spremenljivko (*c*) in rezultat seštevanja izpiše.

```
1 def sestej(a, b):
2     c = a + b
3     print(c)
4     # tale komentar je še del funkcije
5 # tale komentar ni več del funkcije
```

Kaj pa se zgodi, ko program s tole definicijo poženemo. Navidez se ne zgodi nič, če pa v ukazno vrstico napišemo ime pravkar definirane funkcije, bi moral Python izpisati nekaj podobnega temu:

```
<function sestej at 0x000001C24E1481E0>
```

Kaj to pomeni? To pomeni, da se je v našem *imenskem prostoru* (pojem bomo razložili v kratkem) pojavilo ime **sestej**, ki ima v ozadju funkcijo, ta pa je shranjena nekje v pomnilniku (natančneje na pomnilniškem naslovu 0x000001C24E1481E0). Ko smo izvedli zgornjo kodo, smo torej dobili definicijo funkcije **sestej**, ki jo zdaj lahko pokličemo.

Do zdaj smo funkcije vedno klicali tako, da so imenu funkcije sledili oklepaji, znotraj katerih smo našli vrednosti argumentov, nad katerimi smo želeli funkcijo poklicati. In seveda je tako tudi v primeru funkcij, ki jih definiramo sami. Če bi torej želeli izpisati vsoto števil 5 in 7, bi lahko izvedli klic

```
>>>sestej(5,7)
12
```

Ko smo funkcijo definirali se torej koda znotraj funkcije sploh ni izvedla. Izvedla se je zgolj njena definicija, ki nam je njeno ime umestila v imenski prostor (podobno,

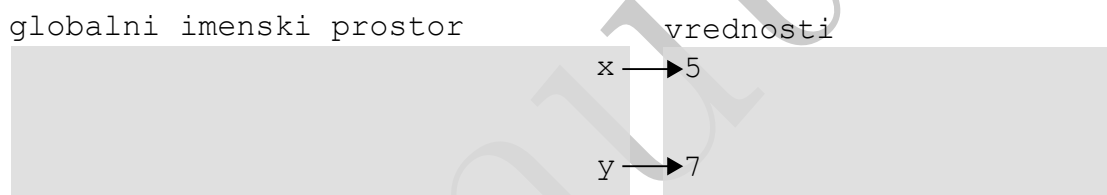
kot če smo nekemu imenu – spremenljivki, priredili neko vrednost). Dejanska izvedba stavkov znotraj funkcije pa se je izvršila šele, ko smo funkcijo poklicali. Mimogrede, če bi v funkciji imeli kakšno sintaktično napako, kot je npr. uporaba nedefinirane spremenljivke, bi jo Python našel šele ob klicu funkcije.

7.3 Globalni imenski prostor

Vsakič, ko v Pythonu definiramo novo spremenljivko, se ime, preko katerega bomo dostopali do vrednosti te spremenljivke, shrani v t.i. *imenski prostor*. Podobno se zgodi ob definiciji funkcije, le da se v tem primeru za imenom funkcije skriva vsebina funkcije, ki se bo izvedla, ko jo bomo poklicali. Ko npr. definiramo spremenljivki `x` in `y` z uporabo kode

```
>>> x = 5
>>> y = 7
```

se v imenskem prostoru pojavita imeni `x` in `y`, za katerimi se skrivata podani vrednosti, kot prikazuje slika 7.1. Preko imen `x` in `y` lahko zdaj dostopamo do



Slika 7.1 Imena v imenskem prostoru kažejo na konkretne vrednosti v pomnilniku.

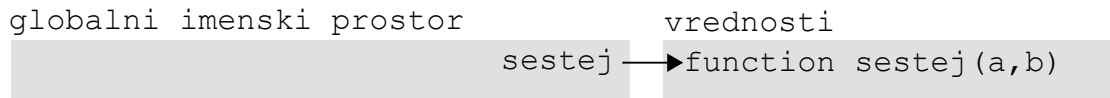
vrednosti, ki se skrivajo v ozadju, ne da bi se morali zavedati kje konkretno v pomnilniku so te vrednosti shranjene, kar nam bistveno olajša življenje.

Definicije novih imen, pa naj gre za imena spremenljivk ali funkcij, ki jih ustvarimo izven funkcij, se shranijo v t.i. *globalni imenski prostor*. Zato tem imenom pogosto rečemo kar globalna imena, spremenljivkam pa *globalne spremenljivke*. Če obstaja globalni imenski prostor pa bo verjetno obstajal tudi lokalni. Poglejmo si, kaj se zgodi, ko funkcijo pokličemo.

7.4 Kaj se zgodi ob klicu funkcije in lokalni imenski prostor

Kot smo že omenili, se ob definiciji funkcije v globalnem imenskem prostoru ustvari novo ime, ki je enako imenu funkcije. To kaže na samo funkcijo, tako da bomo lahko le-to kasneje preko imena tudi poklicali. Situacijo po definiciji funkcije `sestej` prikazuje slika 7.2.

Dopolnimo program, v katerem smo napisali funkcijo `sestej`, še z njenim klicem.



Slika 7.2 Ob definiciji funkcije v imenskem prostoru dobimo novo ime, ki je enako imenu funkcije. Za tem imenom se skriva naša funkcija.

```

1 def sestej(a, b): # definicija funkcije
2     c = a + b
3     print(c)
4 x = 5
5 y = 7
6 sestej(x,y) # klic funkcije

```

Vrstice programa od 1, 4 in 5 bi morale biti zdaj že popolnoma jasne. Kaj pa se zgodi, ko program pride do vrstice 6? Ustvari se lokalni imenski prostor funkcije **sestej**, znotraj katerega bo funkcija ustvarila svoje lokalne spremenljivke. V lokalnem imenskem prostoru se najprej ustvarita lokalni spremenljivki z imeni **a** in **b**, ki predstavljata *plitvi* kopiji spremenljivk **x** in **y**, tj. spremenljivk, s katerimi smo funkcijo poklicali. Enako posledico bi imela prireditve

```

>>> a = x
>>> b = y

```

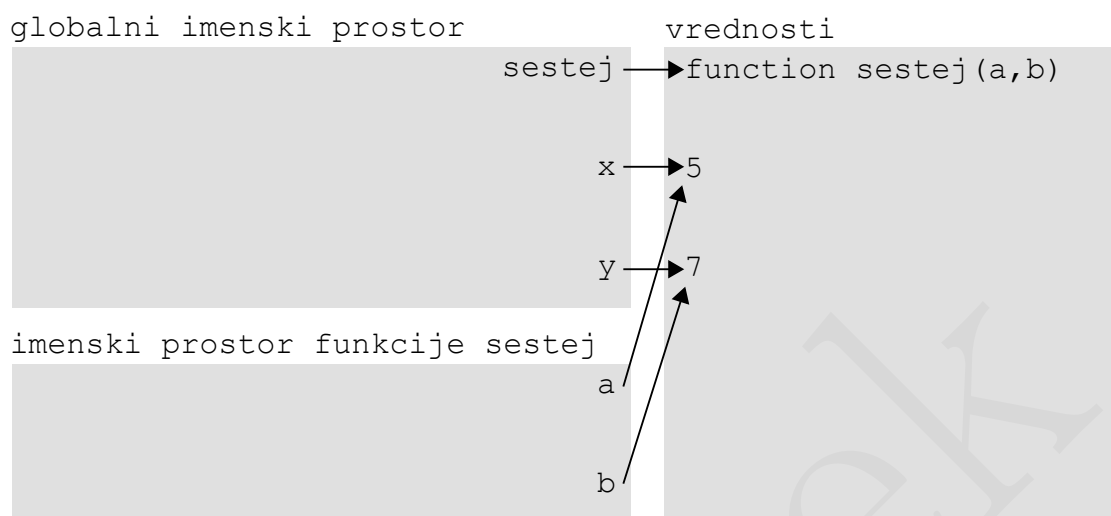
s to razliko, da bi se imeni **a** in **b** ustvarili v globalnem imenskem prostoru.

Plitva kopija pomeni, da vrednost, ki je shranjena v pomnilniku dobi dodatno ime, brez da bi se dejansko kopirala (to bi bila globoka kopija), s čimer smo s pomnilniškim prostorom veliko bolj varčni. O tem bomo še govorili, zaenkrat pa se vrnimo k naši funkciji in njenem lokalnem imenskem prostoru. Situacijo ob klicu funkcije prikazuje slika 7.3.

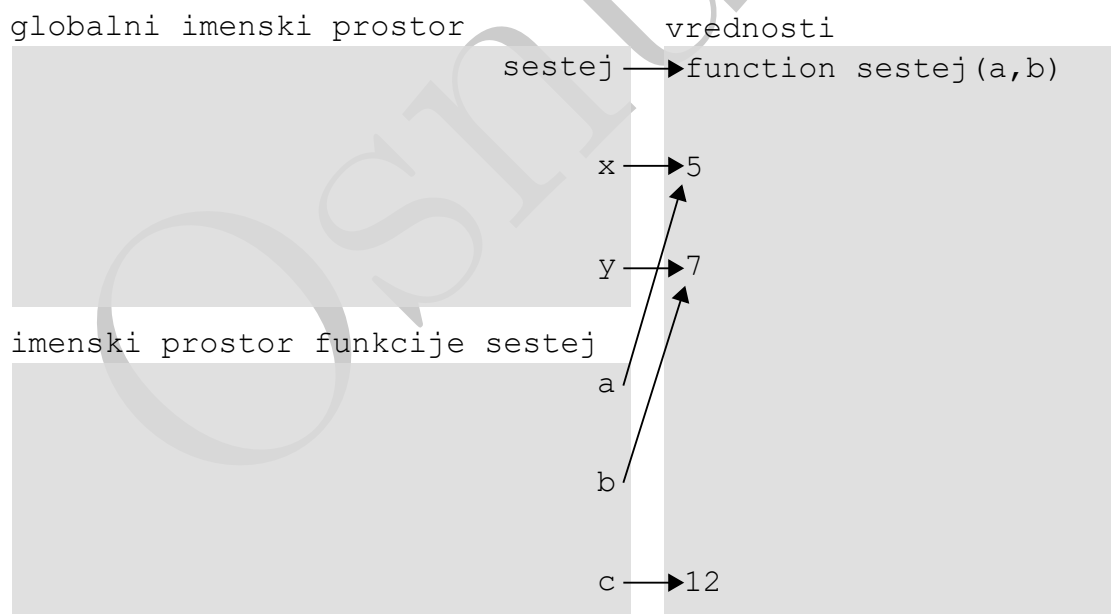
Vsa imena, ki jih bomo v nadaljevanju definirali znotraj funkcije, bodo ustvarjena v lokalnem imenskem prostoru funkcije. Ko naš program na primer izvede vrstico 2 (ta se je ob definiciji funkcije preskočila in se izvede šele ob njenem klicu), bo prišlo do situacije, kot jo prikazuje slika 7.4

Kaj pa bi se zgodilo, če bi znotraj funkcije definirali ime, ki obstaja že v globalnem imenskem prostoru. Nič posebnega. Spremenljivka s tem imenom bi se ustvarila v lokalnem imenskem prostoru funkcije in to na globalno spremenljivko ne bi vplivalo. Zgodilo bi se nekaj takega kot prikazuje slika 7.5.

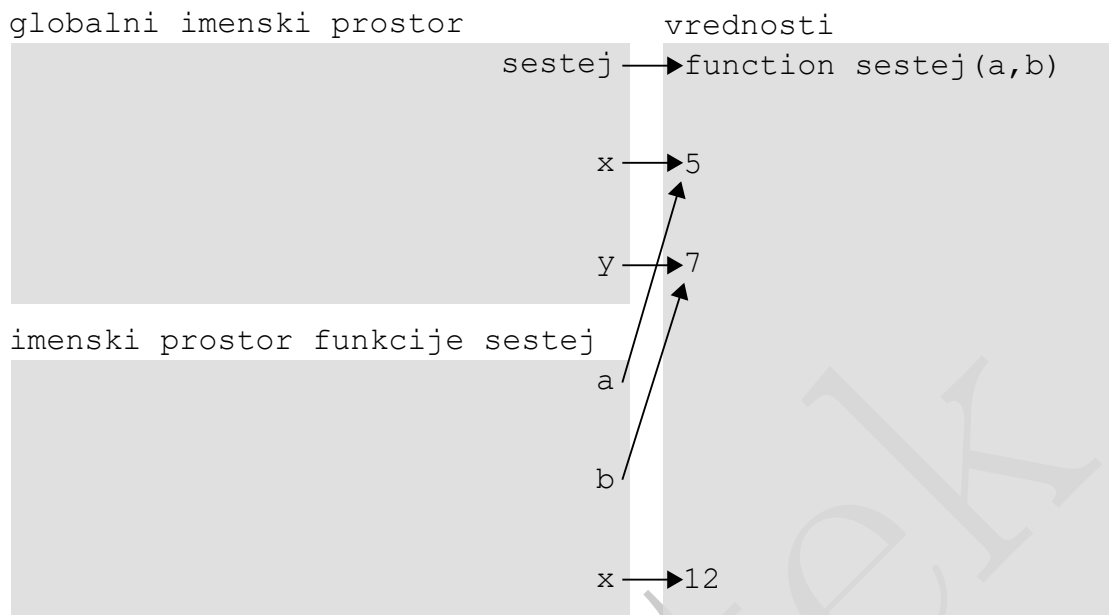
Zakaj je tak način delovanja dober? Če bi morali znotraj funkcij paziti, da ne uporabljamo enakih imen, kot so že definirana izven funkcij, potem bi morali že vnaprej predvideti kakšna imena bodo pri programiranju uporabljali vsi bodoči uporabniki naših funkcij. Prav tako bi morali biti zelo pazljivi, ko bi obstoječe funkcije uporabljali mi. Vedeti bi morali katere spremenljivke za izpis nečesa na



Slika 7.3 Ob klicu funkcije se ustvari njen lokalni imenski prostor znotraj katerega se dodatno ustvariijo plitve kopije vrednosti, s katerimi smo funkcijo poklicali.



Slika 7.4 Vsa imena, ki jih definiramo znotraj funkcije, se ustvariijo zgolj v lokalnem imenskem prostoru te funkcije.

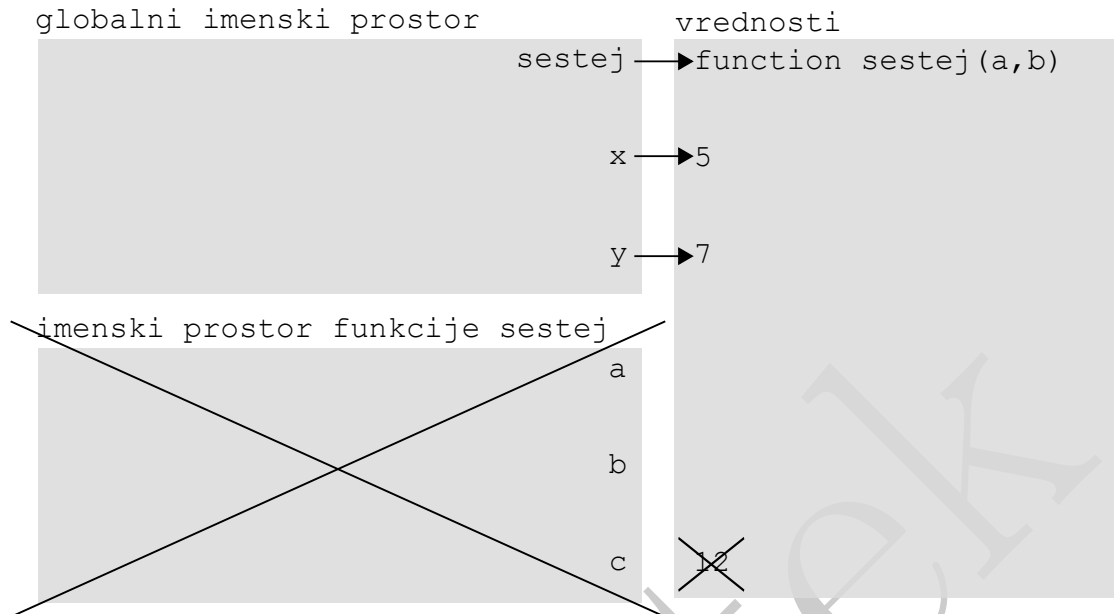


Slika 7.5 Znotraj funkcije lahko uporabljamo enaka imena spremenljivk kot izven funkcije in s tem ne vplivamo na globalne spremenljivke.

zaslon na primer uporablja funkcija `print`. Tem imenom bi se morali izogibati, kar pa bi bilo skrajno nerodno in nesmiselno.

Vprašanje, na katerega moramo še odgovoriti je, kaj se zgodi, ko se funkcija izvede do konca. V našem primeru se funkcija konča po izpisu vrednosti spremenljivke `c` (vrstica 3). Ko se funkcija konča, njenega lokalnega imenskega prostora ne potrebujemo več. Če bomo funkcijo še enkrat poklicali, bo Python pač ustvaril nov lokalni imenski prostor. Iz tega razloga po končanju izvedbe funkcije lokalni imenski prostor funkcije izgine. V našem konkretnem primeru torej imena `a`, `b` in `c` izginejo. Kaj pa vrednosti? Do vrednosti 12 ne moremo več dostopati preko nobene spremenljivke, zato se lahko izbriše tudi ta. Vrednosti 5 in 7 po drugi strani ostaneta, saj nanju še vedno kažeta imeni `x` in `y`. To prikazuje slika 7.6.

Iz globalnega imenskega prostora do lokalnih imenskih prostorov uporabljenih funkcij torej ne moremo dostopati, saj se po zaključku izvajanja funkcij (ko izvedba programa preide spet v globalni imenski prostor), lokalni imenski prostori izbrišejo. Kaj pa obratno? Iz lokalnega imenskega prostora funkcije, lahko dostopamo do globalnega (tudi zato se mu reče globalni), kar pomeni, da lahko dostopamo do vrednosti globalnih spremenljivk. Še pomembneje pa je to, da lahko iz lokalnega imenskega prostora funkcij, dostopamo do imen globalno definiranih funkcij. To pomeni, da lahko iz posamezne funkcije pokličemo druge funkcije (gnezdenje funkcij) ali pa tudi samo sebe. Slednjemu se reče *rekurzija*, ampak pustimo to za kdaj drugič.



Slika 7.6 Po izvedbi klica funkcije, se njen imenski prostor izbriše.

Napišimo malo razširjen program, ki bo seštel vrednosti dveh seznamov. Pri tem si bomo pomagali z definicijo dveh funkcij.

```

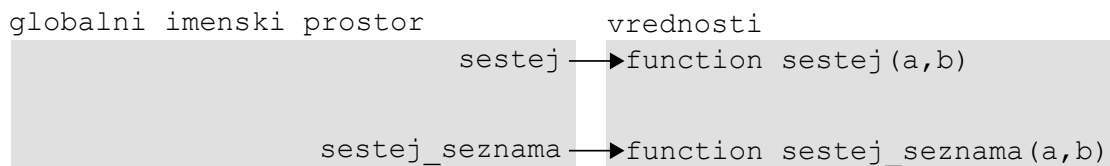
1 def sestej(a, b): # seštej in izpiši
2     c = a + b
3     print(c)
4
5 def sestej_seznam(a,b): # seštej istoležne elemente
6     for i in range(len(a)):
7         sestej(a[i], b[i])
8
9 sestej_seznam([1,2,3],[4,5,6]) # klic funkcije

```

Iz funkcije `sestej_seznam` torej kličemo funkcijo `sestej`. Ali je to dovoljeno? Seveda. Imeni `sestej` in `sestej_seznam` bomo po izvedbi vrstic 1 in 5 imeli v globalnem imenskem prostoru, kot prikazuje slika 7.7.

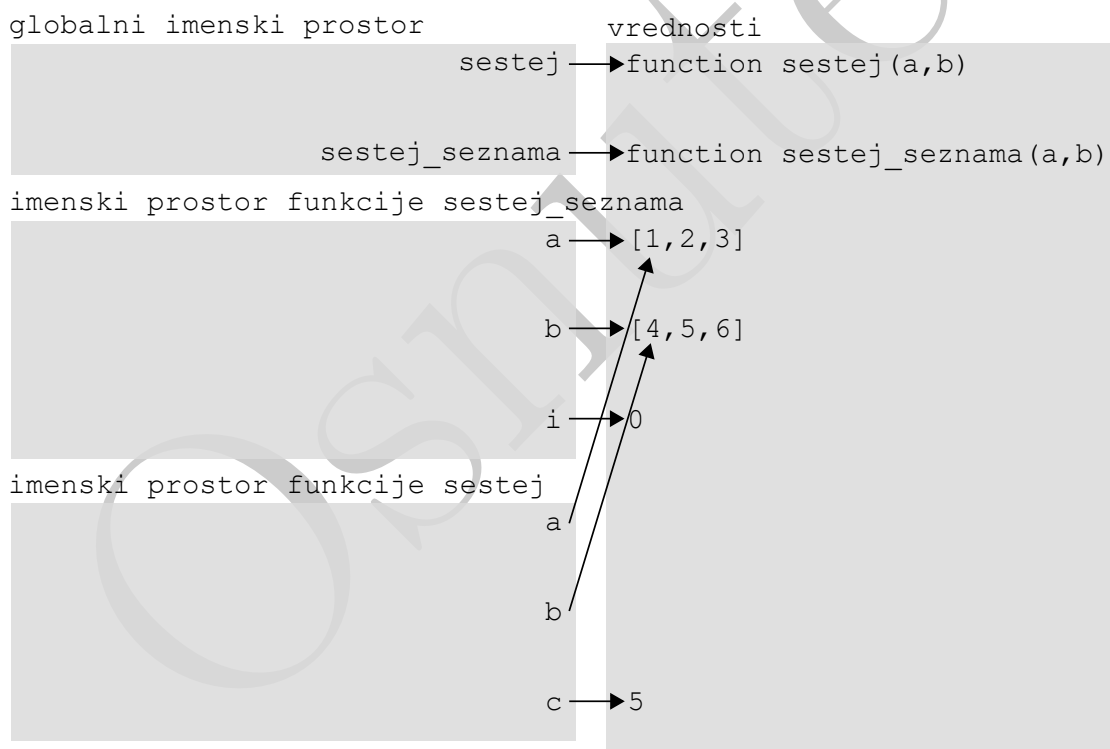
Ker je globalni imenski prostor viden tudi iz lokalnih imenskih prostorov posameznih funkcij, jih lahko od tam tudi pokličemo. V sled temu je zgornji program popolnoma pravilen.

Če program pogledamo podrobneje, lahko vidimo, da obe funkciji uporabljata enaka imena spremenljivk. Tudi to ne bo povzročalo nobenih težav, saj bo vsaka funkcija dobila svoj lasten lokalni imenski prostor. Ko bomo poklicali funkcijo `sestej_seznam`, bo ta dobila lokalni imenski prostor. Ko bomo iz te funkcije poklicali funkcijo `sestej`, bo ta dobila svoj imenski prostor, ki se s prostorom



Slika 7.7 Imeni definiranih funkcij sta shranjeni v globalnem imenskem prostoru, zato jih lahko pokličemo od kjerkoli.

funkcije `sestej_seznam` ne bo prekrival. Lokalne imenske prostore si torej lahko predstavljamo kot ločene mehurčke, ki se med seboj ne prekrivajo. Stanje našega programa ob prvi izvedbi funkcije `sestej` do vključno vrstice 2 prikazuje slika 7.8.



Slika 7.8 Lokalni imenski prostori funkcij so med seboj ločeni.

Vprašanje za razmislek – zakaj se znotraj funkcije `sestej` ne ustvarita novi vrednosti, na kateri bosta kazali imeni `a` in `b`?

7.5 Vsaka funkcija vrača rezultat

V splošnem pri programiranju ločimo dva tipa funkcij, in sicer tiste, ki nekaj uporabnega vrnejo in tiste, ki nekaj uporabnega naredijo (vrnejo pa nič). V določenih programskih jezikih ti dve skupini nosijo celo posebna imena in sta tudi drugače definirani. Kaj pa v jeziku Python? V skupino funkcij, ki nekaj uporabnega vračajo bi lahko uvrstili npr. funkcijo `input`, ki prebere uporabnikov vnos in tega vrne kot podatkovni tip `str`. V skupino funkcij, ki ne vračajo nič kaj preveč uporabnega, je pa uporabno tisto, kar naredijo, pa spada funkcija `print`. Dejstvo je, da v Pythonu vsaka funkcija nekaj vrne pa tudi, če to ni čisto nič uporabnega. Poglejmo si kaj vrne funkcija `print`. Kako? Rezultat funkcije `print` bomo shranili v spremenljivko in vrednost te spremenljivke izpisali.

```
>>>a = print("testni izpis")
testni izpis
>>>print(a)
None
```

Kaj se torej skriva v rezultatu funkcije `print`? Dobesedno nič oziroma `None`. Funkcija nekaj vrne, in sicer vrne nič. Preverimo lahko tudi njegov podatkovni tip.

```
>>>print(type(a))
<class 'NoneType'>
```

Nič oziroma `None` je torej poseben podatek, ki pripada podatkovnemu tipu nič oziroma `NoneType`. Ni sicer veliko, ampak nekaj pa je. Enak rezultat vračajo funkcije, ki smo jih definirali v prejšnjem razdelku. Lahko preverite sami.

Kaj pa če bi želeli, da naša funkcija vrne nekaj uporabnega? V tem primeru moramo od nje to eksplicitno zahtevati, in sicer s stavkom `return`.

Spremenimo funkcijo `seštej`, tako da bo vsoto dveh števil vračala in ne izpisovala.

```
1 def seštej(a, b): # seštej in vrni
2     c = a + b
3     return c
```

Z uporabo stavka `return` smo torej povedali, da želimo, da naša funkcija vrne vrednost spremenljivke `c`. Ali nismo tega naredili že prej? Ne. V prejšnji različici je funkcija vrednost spremenljivke `c` zgolj izpisovala. Ko se je funkcija končala, je njen lokalni imenski prostor izginil in z njim tudi vrednost spremenljivke `c`. Pogosto pa želimo rezultate funkcij uporabiti tudi v drugih delih naših programov (npr. ko uporabljamo funkcijo `input` želimo z uporabnikovim vnosom ponavadi nekaj uporabnega narediti in ga ne zgolj izpisati na zaslon). To lahko dosežemo s stavkom `return`. Kaj se zgodi, če funkcijo v našem programu zdaj še pokličemo. Razširimo program na sledeč način.

```
1 def seštej(a, b): # seštej in vrni
```

```

2     c = a + b
3     return c
4 sestej(4,5)

```

Program tokrat ne izpiše ničesar. Zakaj ne? Ker tega od njega nismo nikjer zahtevali. Kaj torej naredi klic funkcije `sestej`. V konkretnem primeru nič uporabnega, saj izračuna vsoto števil 4 in 5, rezultat shrani v spremenljivko `c` in ko se funkcija zaključi, le-ta izgine, saj nanj ne kaže nobeno ime več. Kako pa bi lahko dobljeno vrednost uporabili še kje druge v našem programu? Podobno kot pri uporabi funkcije `input` – tako, da bi rezultat funkcije priredili spremenljivki.

```

1 def sestej(a, b): # seštej in vrni
2     c = a + b
3     return c
4 rezultat = sestej(4,5)
5 print(rezultat)

```

V zgornjem primeru bomo rezultat izpisali, lahko pa bi z njim naredili tudi karkoli drugega.

Stavek `return` ima dvojno vlogo. Ob njegovem klicu funkcija vrne rezultat, poleg tega pa se njeno izvajanje prekine (podobno, kot če uporabimo stavek `break` v kombinaciji z zanko `while` ali `for`).

Povadimo zdaj to na iskanju praštevil. Najprej poskusimo napisati funkcijo, ki uporabniku informacijo o tem, ali število je praštevilo ali ne, zgolj izpiše.

Zgled 26 *Napiši funkcijo, ki kot argument prejme celo število in izpiše, če je podano število praštevilo ali ne.*

Rešitev 26 *Število je praštevilo, če tega ne deli nobeno od njega manjše naravno število, ki je večje od 1. Sprehoditi se moramo torej čez interval od 2 do našega števila – 1 in za vsako število iz intervala preveriti, če deli naše število. Če na intervalu najdemo vsaj enega delitelja, število ni praštevilo in lahko iskanje takoj prekinemo. Če delitelja na celotnem intervalu ne najdemo, lahko sklepamo, da je število praštevilo.*

```

1 def prastevilo(stevilo):
2     for i in range(2, stevilo): # razpon preiskovanja
3         if stevilo % i == 0: # delitelj?
4             print(stevilo, "ni praštevilo")
5             break # dovolj je, da najdemo enega delitelja
6     else: # če se je zanka odvirtela do konca
7         print(stevilo, "je praštevilo")

```

Poskusimo zdaj funkcijo spremeniti, tako da ne bo ničesar izpisovala, ampak bo uporabniku podala povratno informacijo o tem, če je število praštevilo ali ne.

Zgled 27 *Napiši funkcijo, ki kot argument prejme celo število in vrne vrednost **True**, če je to število praštevilo, sicer pa vrne vrednost **False**.*

Rešitev 27

```
1 def prastevilo(stevilo):
2     for i in range(2, stevilo): # razpon preiskovanja
3         if stevilo % i == 0:
4             return False # prekine funkcijo in vrne False
5     return True # for se je odvrtil do konca
```

Ta rešitev je bistveno lepša in enostavnejša. Iz nje vidimo dodatno prednost stavka **return**, ki poleg vračanja rezultata prekine izvajanje funkcije. Ko smo v zanki **for** našli prvega delitelja, smo prekinili izvajanje funkcije in vrnili rezultat **False**. S prekinitvijo izvajanja funkcije se je prekinila tudi zanka, zato **break** ni več potreben. Če je program prišel do vrstice številka 5, zagotovo nismo našli nobenega delitelja, saj bi sicer funkcija že vrnila **False** in se nehala izvajati. Zato lahko v vrstici 5 brezpogojno vrnemo vrednost **True**. Tako tukaj ne potrebujemo niti stavka **if**. Dodatna prednost te rešitve je tudi to, da lahko zdaj rezultat preverjanja uporabimo tudi kje drugje. Lahko na primer napišemo funkcijo, ki izpiše vsa praštevila v določenem razponu.

Zgled 28 *Napiši funkcijo, ki kot argument prejme celo število in izpiše vsa praštevila do vključno podanega števila.*

Rešitev 28 *Vse kar je potrebno narediti je sprehod čez interval števil od 2 do podanega števila, klic in preverjanje rezultata klika funkcije, ki smo jo definirali zgoraj.*

```
1 def prastevila(stevilo):
2     for kandidat in range(2, stevilo+1): # kandidati
3         if prastevilo(kandidat): # ali je praštevilo
4             print(kandidat)
```

Rešitev je izjemno enostavna. Sprehodili smo se čez vse možne kandidate za praštevila in za vsakega preverili, če je praštevilo. Kako? Tako, da smo poklicali funkcijo, ki vrne **True**, če je podano število praštevilo. Klic te funkcije smo vstavili v stavek **if**, ki je izpisal število v primeru izpolnjenosti pogoja.

7.6 Izbirni argumenti

V določenih primerih želimo, da imajo določeni argumenti funkcije svoje vrednosti že vnaprej določene, razen v primeru, da uporabnik želi za te argumente uporabiti

druge vrednosti. Če torej uporabnik vrednosti argumentov ne bo podal, bodo uporabljene njihove privzete vrednosti. V nasprotnem primeru bodo uporabljene uporabnikove vrednosti. Tak primer uporabe funkcij smo srečali že pri funkciji `print`, ki ima kar nekaj izbirnih argumentov. Privzeto gre funkcija `print` po vsakem klicu v novo vrstico (argument `end` je privzeto enak znaku za novo vrstico

`\n`), v primeru več podanih vrednosti pa te izpišejo tako, da se med njih vrine presledke (argument `sep` je privzeto enak presledku). Njune privzete vrednosti lahko povežemo, tako da jih specificiramo ob klicu, npr. kot

```
>>> print(1,2,3,sep=' ',end=' ')
1+2+3
```

Podobno lahko specificiramo izbirne argumente in njihove vrednosti pri definiciji svojih funkcij.

```
def ime_funkcije(arg1, arg2, ..., opc1=v1, opc2=v2, ...):
```

Paziti moramo samo na to, da so tisti argumenti, ki nimajo privzetih vrednosti vedno podani pred tistimi, ki privzete vrednosti imajo.

Povadimo to na malo bolj splošnem Fibonaccijevem zaporedju. Najprej poskusimo brez uporabe opsijskih argumentov.

Zgled 29 *Napiši funkcijo, ki vrne Fibonaccijevo zaporedje števil, pri čemer naj uporabik poda dolžino zaporedja in prvi dve števili v zaporedju.*

Rešitev 29 *Fibonaccijevo zaporedje je zaporedje števil, v katerem sta prva dva člena enaka številu 1, vsak nadaljnji člen pa je enak vsoti prejšnjih dveh členov. V funkciji želimo generirati bolj splošno Fibonaccijevo zaporedje, ki se začne s poljubnima številoma, npr. a in b . Znotraj funkcije bomo najprej preverili, če je željena dolžina zaporedja n enaka 1 (v tem primeru vrnemo seznam, ki vsebuje zgolj prvo podano število $[a]$). Sicer bomo naredili seznam z obema podanima številoma $[a, b]$. Potem bomo še $n-2$ -krat izvedli zanko, v kateri bomo v seznam vsakič dodali nov element, ki bo predstavljal vsoto trenutno zadnjih dveh elementov seznama.*

```
1 def fibonacci(n, a, b):
2     if n == 1:
3         return [a]
4     f = [a, b]
5     for i in range(3, n+1):
6         f.append(f[-1]+f[-2])
7     return f
```

V primeru, da uporabnik želi imeti zaporedje dolžine 1, funkcija vrne zaporedje z enim elementom, in sicer `a`. V nasprotnem primeru naredi začetno zaporedje z elementoma `a` in `b` in potem v zanki doda ustrezno število dodatnih elementov, ki vsakič predstavljajo vsoto zadnjih dveh elementov zaporedja. S to rešitvijo sicer ni nič narobe, je pa dejstvo to, da si kot Fibonaccijevo zaporedje ponavadi predstavljamo zaporedje števil, ki se začne z vrednostma 1, 1. Smiselno bi torej bilo, da se privzeto naše zaporedje začne s števili 1, 1, razen če uporabnik tega ne specificira drugače.

Zgled 30 *Napiši funkcijo, vrne Fibonaccijevo zaporedje števil, pri čemer naj uporabnik poda dolžino zaporedja. Uporabnik lahko poda tudi prvi dve števili v zaporedju, ki sta privzeto enaki 1.*

Rešitev 30 *Funkcijo bomo dopolnili tako, da sta argumenta `a` in `b` opcijska in da sta privzeto postavljena na vrednost 1.*

```

1 def fibonacci(n, a=1, b=1):
2     if n == 1:
3         return [a]
4     f = [a, b]
5     for i in range(3, n+1):
6         f.append(f[-1]+f[-2])
7     return f

```

Zgornjo funkcijo lahko torej pokličemo tudi tako, da podamo samo dolžino zaporedja. V tem primeru bosta prvi dve števili v zaporedju enaki 1, 1. V primeru, da jo bomo poklicali tako, da podamo še vrednosti za argumenta `a` in `b` pa bosta za začetna elementa uporabljeni ti vrednosti.

Osnutek

8 Uporaba in pisanje modulov

8.1 Kaj so moduli?

Moduli predstavljajo Pythonove datoteke, ki vsebujejo implementacijo določenih funkcij, spremenljivk in razredov (angl. *classes*). Module lahko vključimo v svoje programe in na ta način razširimo osnovne funkcionalnosti jezika Python. Primeri že vgrajenih modulov, ki jih ni potrebno posebej namestiti, so modul `math`, v katerem so definirane določene matematične funkcije in konstante, modul `time` za vračanje podatkov o času in tvorjenje zakasnitev ter modul `random` za delo s (psevdo)naključnimi števili.

8.2 Uporaba modulov

Module lahko v svoje programe vključimo na različne načine, v vseh primerih pa uporabljamo rezervirano besedo `import`. Če želimo npr. v naš program uvoziti celoten modul `math`, lahko to naredimo s sledečo vrstico:

```
import math
```

oziroma v splošnem

```
import ime_modula
```

Najprej lahko preverimo kaj uvožen modul dejansko ponuja. Če je pisec modula bil priden in napisal tudi dokumentacijo (v obliki komentarjev), bomo za to lahko uporabili funkcijo `help`

```
help(ime_modula)
```

Funkcija nam bo izpisala nekaj osnovnih informacij o modulu, tako da bo uporaba lažja. Seveda pa lahko informacije o modulu poiščemo tudi na internetu, ki pa včasih ni na voljo (npr. v času pisanja kolokvijev in izpitov), zato se je dobro navaditi tudi uporabe zgoraj omenjene funkcije.

Če bi zdaj želeli dostopati do posamezne funkcije, ki je v uvoženem modulu definirana, bi to naredili na sledeč način

```
ime_modula.ime_funkcije(argumenti)
```

Če bi npr. želeli izračunati sinus števila shranjenega v spremenljivki `x` in rezultat shraniti v spremenljivko `y`, bi za to uporabil funkcijo `sin`, ki je vsebovana v modulu `math`. Poklicali bi jo takole

```
y=math.sin(x)
```

Včasih imajo moduli zelo dolga in težko berljiva imena. Če želimo modul uvoziti pod drugačnim imenom (lahko bi rekli psevdonimom), uvoz dopolnimo z `as` stavkom:

```
import dolgo_ime_modula as psevdonim
```

Tako lahko pri klicanju funkcij (ali pa česarkoli že) modula podajamo le kratko ime modula. V prejšnjem primeru bi kodo lahko spremenili na sledeč način:

```
import math as m
y=m.sin(x)
```

V določenih primerih pa želimo iz modula uvoziti le določeno funkcijo (spremenljivko, razred). Takrat lahko uporabimo rezervirano besedo `from`, in sicer takole

```
from ime_modula import ime_funkcije
```

S takim načinom uvažanja smo uvozili le tisto kar potrebujemo, poleg tega pa zdaj pri klicu funkcije imena modula ni potrebno več podajati. Primer sinusa bi se spremenil v sledečo kodo

```
from math import sin
y=sin(x)
```

Zdaj smo iz modula uvozili zgolj funkcijo `sin` – če bi želeli imeti še kakšno drugo funkcijo, npr. kosinus, bi jo morali uvoziti ločeno oziroma hkrati s funkcijo `sin`. To bi naredili takole:

```
from math import sin,cos
```

Lahko pa naredimo še nekaj, kar ponavadi ni priporočljivo. Uvozimo lahko vse, kar je v modulu definirano, in sicer namesto imena funkcije podamo `*`, ki se v računalništvu velikokrat uporablja kot simbol za *vse*. Rekli bomo torej *iz modula uvozi vse*:

```
from ime_modula import *
```

V primeru modula `math` bomo zapisali takole:

```
from math import *
```

Zakaj tak način uvažanja ni priporočljiv? Vse funkcije, spremenljivke in razrede modula smo zdaj dobili v naš globalni imenski prostor. Pri tem je velika verjetnost, da smo si s tem povozili kakšno od spremenljivk, ki jo tam že uporabljamo in ima enako ime kot kakšna izmed funkcij ali spremenljivk definiranih v modulu. Zato se takemu način uvažanja modulov izogibamo.

8.3 Definicija in uporaba lastnih modulov

Vsi programi, ki smo jih do zdaj napisali, predstavljajo module, ki jih lahko uvozimo v druge programe. To pomeni, da pri reševanju nekega (bolj kompleksnega) problema ni potrebno vse kode napisati v isti datoteki, ampak lahko datoteko razdelimo po smiselnih moduli, pri čemer lahko funkcije prvega modula uporabljamo v drugem in obratno. Pri tem lahko uporabimo kodo opisano v prejšnjem razdelku. Paziti moramo le na to, da se modul, ki ga uvažamo nahaja v isti mapi kot modul, v katerega kodo uvažamo. V nasprotnem primeru moramo pri uvažanju modula do drugega modula podati še pot do njega. Če se modul, ki ga želimo uvoziti, npr. nahaja v podmapi mapi `podmapa`, ga bomo uvozili na sledeč način:

```
import podmapa.ime_modula
```

Če v tem primeru ne želimo, da modul vsakič posebej kličemo z imenom `podmapa.ime_modula`, ga je smiselno uvoziti pod krajšim imenom takole

```
import podmapa.ime_modula as ime_modula
```

Mogoče se sprašujete zakaj poti ni bilo potrebno podajati pri uvažanju modula `math`. Dejstvo je, da so določeni moduli v Python že vgrajeni, sicer pa Python module, poleg v trenutni delovni mapi, išče tudi v mapi `lib\site-packages`, kamor se shranijo namestitve vseh modulov, ki jih bomo v prihodnosti potencialno še namestili.

8.4 Nameščanje novih modulov

Na spletu obstaja veliko modulov, ki so jih razvili programerji pred nami. Te lahko uporabimo, kadar želimo pri reševanju določenega problema uporabiti višjenivojske sestavine. Če želimo npr. narisati graf povprečne mesečne plače v Sloveniji, nam ni potrebno študirati, kako se lotiti kakršnegakoli risanja v jeziku Python, ampak enostavno uporabimo knjižnico (knjižnica ni nič drugega kot zbirka modulov) `matplotlib` in njene funkcije za risanje grafov. Problem, s katerim se srečamo, je, da tovrstne knjižnice oziroma paketi v osnovni različici Pythona še niso nameščeni (razen, če si nismo namestili distribucije Anaconda ¹) Pred uporabo jih moramo torej namestiti. Problem nameščanja tovrstnih knjižnic in paketov je, poleg včasih mukotrpnega procesa iskanja ustreznih namestitvenih datotek in ročne namestitve, tudi v tem, da za svoje delovanje večina knjižnic in paketov uporablja druge knjižnice in pakete, ti spet druge in tako naprej. Temu rečemo odvisnost med paketi (angl. *package dependency*). Da pa se s tem običajnemu uporabniku Pythona ni potrebno ukvarjati, Python okolje vsebuje orodje `pip` (angl. *pip installs packages*), ki preko

¹Anaconda je distribucija Pythona za znanstveno računanje, ki ima nameščenih že večino knjižnic, ki jih za tako računanje potrebujemo. Dostopna je na povezavi <https://www.anaconda.com/>.

repozitorija PyPI (angl. *Python package index*) poišče in namesti ustrezne pakete avtomatsko. Nameščen je že skuppaj z osnovno distribucijo Pythona. Vse kar poleg tega potrebujemo je še internetna povezava in ime knjižnice oziroma paketa, ki ga želimo namestiti. Orodje `pip` bomo pognali iz systemske ukazne vrstice (v operacijskem sistemu Windows jo zaženemo tako, da v start meni vpišemo `cmd`). V primeru, da smo ob namestitvi Pythona obkljukali opcijo *Add Python to path*, lahko orodje `pip` poženemo iz poljubne lokacije. V nasprotnem primeru se moramo premakniti v mapo, kjer `pip` nameščen (podmapa `Scripts` mape, kjer je nameščen Python). Paket z imenom `ime_paketa` zdaj namestimo s sledečim ukazom

```
> pip install ime_paketa
```

in `pip` bo poskrbel za vse ostalo.

9 Spremenljivost podatkovnih tipov in terke

9.1 Kaj je spremenljivost?

Določeni podatkovni tipi v Pythonu so spremenljivi (angl. *mutable*), določeni pa ne. Kaj to pomeni? Če je nek podatek nespremenljiv (angl. *immutable*), to pomeni, da ga po tistem, ko je enkrat ustvarjen, ne moremo več spreminjati. Lahko pa naredimo nov podatek, ki odraža spremembo, ki jo želimo nad podatkom narediti. Primeri nespremenljivih podatkovnih tipov so števila tipa `int` in `float`, niz oziroma `str` in `bool` (spremenljivost osnovnih podatkovnih tipov v jeziku Python prikazuje tabela 9.1). To so torej skoraj vsi podatkovni tipi, ki smo jih do sedaj spoznali. Če je določen podatek spremenljiv, potem ga lahko spreminjamo tudi kasneje. Primer spremenljivega podatkovnega tipa je seznam oziroma `list`. Spremenljivost podatkovnih tipov na videz izgleda kot nekaj, s čimer se nam pri osnovah programiranja niti ne bi bilo potrebno ukvarjati. Žal pa ima veliko posledic, ki jih brez razumevanja spremenljivosti težko razumemo, zato je smiselno, da si celoten koncept podrobneje pogledamo.

Tabela 9.1 Spremenljivost osnovnih podatkovnih tipov v jeziku Python

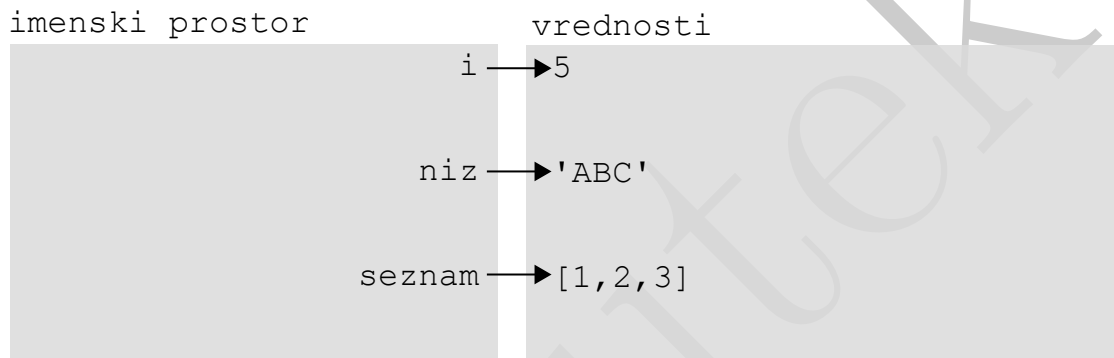
podatkovni tip	opis	spremenljiv
<code>bool</code>	<i>boolean</i> (<code>True</code> , <code>False</code>)	Ne
<code>int</code>	<i>integer</i> (celo število)	Ne
<code>float</code>	<i>floating-point</i> (decimalno število)	Ne
<code>str</code>	<i>string</i> (niz)	Ne
<code>list</code>	<i>list</i> (seznam)	Da
<code>tuple</code>	<i>tuple</i> (terka)	Ne
<code>dict</code>	<i>dictionary</i> (slovar)	Da
<code>set</code>	<i>set</i> (množica)	Da
<code>frozenset</code>	<i>frozenset</i> (nespremenljiva množica)	Ne

9.2 Kaj se zgodi ob prirejanju spremenljivk?

Kaj se zgodi, ko spremenljivki priredimo neko vrednost že vemo. V imenskem prostoru, kjer spremenljivko definiramo, se pojavijo imena, ki smo jih dodelili spremenljivkam, v pomnilniku pa se ustvarijo vrednosti, na katere ta imena kažejo. Zaporedje prireditvenih stavkov

```
>>> i = 1
>>> niz = 'ABC'
>>> seznam = [1,2,3]
```

lahko ponazorimo s sliko 9.1. Kaj pa se zgodi, če spremenljivko priredimo drugi imenski prostor



Slika 9.1 Ob prireditvi se imenu spremenljivke priredi podana vrednost.

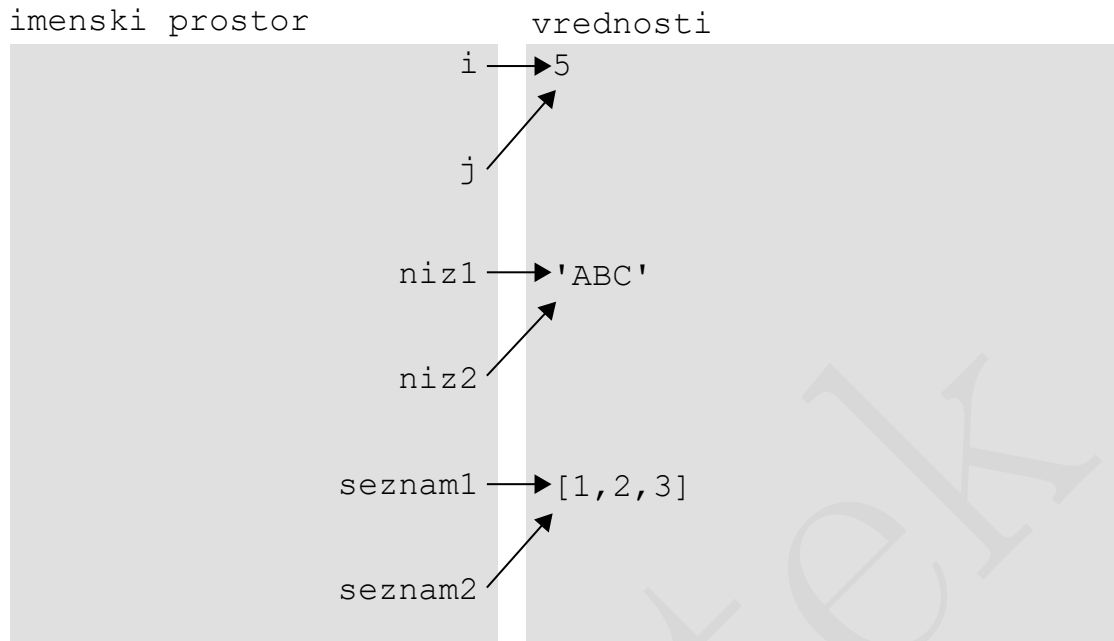
spremenljivki, na primer takole:

```
>>> i = 1
>>> j = i
>>> niz1 = 'ABC'
>>> niz2 = niz1
>>> seznam1 = [1,2,3]
>>> seznam2 = seznam1
```

Nekaj podobnega smo srečali že pri klicu funkcije. Spomnimo se, da Python v takem primeru naredi t.i. *plitvo kopijo* spremenljivke. To pomeni, da vrednost v pomnilniku dobi novo ime. Do dejanskega (*globokega*) kopiranja vrednosti v tem primeru ne pride. To lahko ponazorimo s sliko 9.2. Na tak način je delovanje tako s časovnega stališča (hitrost) kot tudi s prostorskega stališča (poraba pomnilnika) bolj varčno.

9.3 Kaj se zgodi ob spreminjanju vrednosti spremenljivk?

Kaj pa se zgodi, če vrednost nove (ali pa stare) spremenljivke spremenimo? Vse skupaj zavisi od tega ali je podatek, ki ga spreminjamo spremenljiv ali ne. Spo-



Slika 9.2 Ob prireditvi spremenljivke drugi spremenljivki se ustvari plitva kopija spremenljivke.

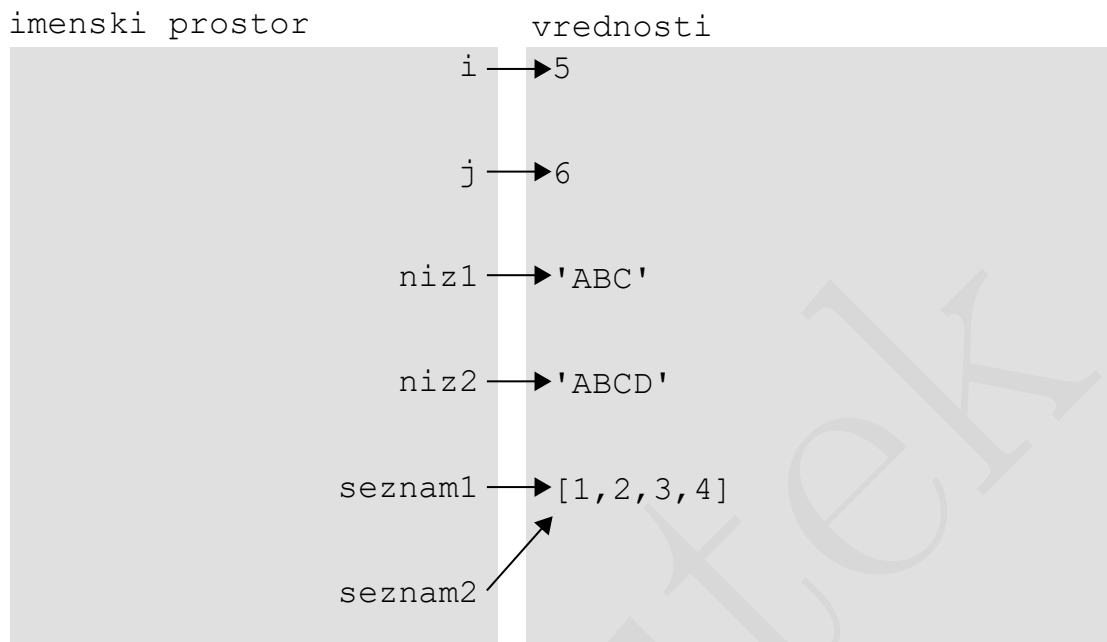
mnimo se. Spremenljiv podatek lahko spreminjamo, ko pa poskusimo spremeniti nespremenljiv podatek, se ustvari njegova kopija (globoka), ki odraža narejeno spremembo. Če npr. uporabimo operator `+=`, bo v primeru spremenljivega podatka spremenjen obstoječ podatek, v primeru nespremenljivega podatka pa bo ustvarjen nov podatek, ki bo odražal narejeno spremembo.

Kaj pa se zgodi v primeru, da na podatek kaže več imen, kot v scenariju zgoraj. Ali se bo po izvedbi spodnje kode sprememba odražala tudi preko drugih imen podatka? Poglejmo si spodnjo kodo.

```
>>> i = 1
>>> j = i
>>> j += 1
>>> niz1 = 'ABC'
>>> niz2 = niz1
>>> niz2 += 'D'
>>> seznam1 = [1, 2, 3]
>>> seznam2 = seznam1
>>> seznam2 += [4]
```

Zanima nas ali se po spreminjanju spremenljivk `j`, `niz2` in `seznam2` spremembe odražajo tudi na spremenljivkah `i`, `niz1` in `seznam1`. Odgovor ni enostaven da ali ne. Odgovor je namreč odvisen od spremenljivosti podatka, ki ga spreminjamo.

Situacijo po spreminjanju podatka z operatorjem `+=` prikazuje slika 9.3. V primeru,



Slika 9.3 Ob spreminjanju spremenljivih podatkov se spremenijo vse plitve kopije podatka.

da je podatek spremenljiv, se torej sprememba odraža na vseh spremenljivkah, ki predstavljajo plitve kopije tega podatka. S tem ko v zgornjem zgledu spreminjamo spremenljivko `seznam2`, spreminjamo tudi spremenljivko `seznam1`. Po drugi strani spreminjanje spremenljivk `j` in `niz2` ustvari globoko kopijo spremenljivk `j` in `niz2`, ki odraža narejeno spremembo. Globoka kopija predstavlja nov podatek, tj. podatek ki se razlikuje od tistega, na katerega kažeta imeni `i` in `niz1`. Posledica tega je, da spreminjanje vrednosti spremenljivk `j` in `niz2` na vrednostih spremenljivk `i` in `niz1` ne vplivajo, saj pripadajo nespremenljivim podatkovnim tipom.

9.4 Ali funkcije spreminjajo vrednosti svojim argumentom?

Spomnimo se, da se ob klicu funkcije ustvari lokalni imenski prostor funkcije. V lokalnem imenskem prostoru se ob klicu spremenljivkam, ki nastopajo kot argumenti funkcije, priredi vrednosti, s katerimi smo funkcijo poklicali. V primeru, da smo funkcijo poklicali z globalnimi spremenljivkami, se argumentom funkcije priredi plitva kopija teh spremenljivk. Vprašanje pa je ali se bodo ob spreminjanju argumentov funkcije spremembe odražale tudi izven funkcije, torej po tem, ko se bo funkcija že končala. Vprašanje lahko ponazorimo s spodnjim zgledom.

Zgled 31 Kakšna je vrednost spremenljivk *st1*, *niz1* in *seznam1* po izvedbi spodnje kode in kakšen bo izpis programa?

```

1 def spremeni(a, b):
2     a += b
3
4     i = 1
5     j = 2
6     spremeni(i, j)
7     print(i)
8
9     niz1 = "ABC"
10    niz2 = "D"
11    spremeni(niz1, niz2)
12    print(niz1)
13
14    seznam1 = [1, 2, 3]
15    seznam2 = [4, 5, 6]
16    spremeni(seznam1, seznam2)
17    print(seznam1)
```

Rešitev 31 Ob klicu funkcije spremenljivki, s katerima funkcijo pokličemo, dobimo kopiji z imeni *a* in *b* v lokalnem imenskem prostoru funkcije. Znotraj funkcije kopijo z imenom *a* spreminjamo. V primeru, da je spremenljivka spremenljivega podatkovnega tipa (npr. seznam) se spreminja obstoječ podatek, na katerega kaže tudi globalna spremenljivka, kar pomeni, da se bo sprememba odražala tudi izven funkcije. V primeru, da je spremenljivka nespremenljivega podatkovnega tipa, se ustvari globoka kopija podatka, ki bo odražala narejeno spremembo. Spremeni se torej zgolj spremenljivka, ki je definirana znotraj funkcije, ta sprememba pa izven funkcije ne bo vidna.

Spremenljivki *i* in *niz1* se torej po klicu funkcije ne bosta spremenili, spremenljivka *seznam1* pa se bo spremenila. Po izvedbi programa bo izpis sledeč:

```

1 # nespremenjena vrednost
ABC # nespremenjena vrednost
[1, 2, 3, 4, 5, 6] # spremenjena vrednost
```

Funkcije torej lahko spreminjajo vrednosti svojim argumentom, tako da so spremembe vidne tudi izven funkcij, ampak samo v primeru, ko so podani argumenti spremenljivega podatkovnega tipa.

9.5 Terke

Zdaj, ko vemo, kaj je to spremenljivost, lahko razložimo tudi, kaj so terke (angl. *tuples*). Terka oziroma `tuple` predstavlja sekvenčen podatkovni tip, ki je nespremenljiv. Ker terke zelo spominjajo na sezname, bi jim lahko rekli tudi nespremenljivi sezname. Načeloma bi lahko pri programiranju shajali tudi brez njih (tako kot bi lahko shajali tudi brez zanke `for`), ampak njihova uporaba v veliko primerih naredi naše programe lepše in boljše (tako kot uporaba zanke `for`).

9.6 Uporaba terk

Terko definiramo z navadnimi oklepaji, tj. `(in)`, znotraj katerih naštejemo elemente. Terko treh elementov, bi lahko definirali na primer takole:

```
>>> terka=("Janez", 1.8, 75)
```

Tudi, če bi elemente našteali brez oklepajev, bi dobili terko. Takole:

```
>>> terka="Janez", 1.8, 75
>>> terka
("Janez", 1.8, 75)
>>> type(terka)
<class 'tuple'>
```

Python je ob naštevanju elementov z vejicami ugotovil, da želimo imeti terko in jo naredil. Python ima nekoliko težav, ko želimo narediti terko dolžine 1, saj si v tem primeru oklepaje razlaga kot operator, ki določa prioriteto. V primeru, da znotraj oklepajev damo zgolj eno npr. celo število, bomo torej dobili podatek, ki pripada podatkovnemu tipu `int` in ne `tuple`:

```
>>> terka=(1)
>>> terka
1
>>> type(terka)
<class 'int'>
```

Ena izmed lastnosti terk je naštevanje elementov, ki jih ločimo z vejicami. Kako v primeru enega elementa povemo, da gre za naštevanje? Tako, da za njim napišemo vejico: `tuple`:

```
>>> terka=(1,)
>>> terka
(1,)
>>> type(terka)
<class 'tuple'>
```

Gre pa seveda tudi brez oklepajev: `tuple`:

```
>>> terka=1,
>>> terka
(1,)
>>> type(terka)
<class 'tuple'>
```

Kaj lahko s terkami počnemo? Podobno kot sezname lahko elemente terke indeksiramo, lahko delamo rezine, lahko preverjamo vsebovanost elementov, z zanko `for` se lahko čez elemente terke sprehajamo itd. Z njimi lahko delamo torej skoraj vse, kar smo delali s sezname. Skoraj vse? Ker so terke nespremenljive, jih seveda ne moremo spreminjati, tako kot lahko spreminjamo sezname. Poskusimo:

```
>>> terka = ("Janez", 1.8, 75)
>>> terka[0] = "Marko"
TypeError: 'tuple' object does not support item assignment
```

Očitno res ne gre. Seveda ne, saj so nespremenljive. Zakaj bi terke potem sploh uporabljali? Nekaj primerov, pri katerih je uporaba terk smiselna, je podanih v nadaljevanju poglavja.

9.7 Seznami terk in razpakiranje elementov terk

Nenapisano pravilo (ki ga seveda lahko kršimo) je, da v sezname shranjujemo homogene podatke, torej podatke, ki se nanašajo npr. na isto spremenljivko. To pomeni, da vsak element seznama obravnavamo na enak način, saj se nanaša na isto količino. Terke se pogosto uporabljajo za shranjevanje heterogenih podatkov, tj. podatkov različnih tipov, ki pa pripadajo isti entiteti, kot je npr. oseba ali meritev. Če si torej želimo pri določeni entiteti zabeležiti več podatkov, lahko uporabimo seznam terk. Na primer, če imamo vzorec oseb, pri čemer za vsako osebo beležimo ime, višino in telesno maso, potem lahko uporabimo seznam terk, pri čemer vsaka izmed terk vsebuje ime, višino in telesno maso dotične osebe. Primer takega seznama bi bil

```
meritve = [("Janez", 1.8, 75),
           ("Ana", 1.65, 60),
           ("Nika", 1.66, 55)]
```

Elementi seznama so torej homogeni, kar pomeni, da bomo vsakega obravnavali enako. Elementi seznama so namreč terke, ki imajo vsakič enako obliko. Na 0-tem indeksu je shranjeno ime osebe, na indeksu 1 višina v metrih in na indeksu 2 telesna masa osebe v kilogramih. Elementi posamezne terke pa očitno pripadajo različnim spremenljivkam.

Tak način predstavitve podatkov bomo srečali velikokrat. Kako pa lahko tako shranjene podatke uporabimo pri nadaljnji analizi. Na primer pri izračunu in

izpisu indeksa telesnih mas posamezne osebe v seznamu. Tako, da se čez seznam sprehodimo z zanko `for` in v vsaki iteraciji zanke tekočo terko *razpakiramo* in tako pridemo do konkretnih vrednosti. To lahko naredimo na sledeč način:

```
for meritev in meritve:
    ime = meritev[0]
    visina = meritev[1]
    masa = meritev[2]
    itm = masa/visina**2
    print("ITM osebe", ime, "je", itm)
```

Do posameznih elementov terke smo torej prišli z njihovim indeksiranjem. Terke pa lahko razpakiramo veliko hitreje, in sicer tako, da terko priredimo drugi terki, ki vsebuje imena spremenljivk, v katere želimo vrednosti shraniti oziroma razpakirati. Takole:

```
(spremenljivka1, spremenljivka2, ...) = terka
```

Paziti moramo le na to, da terka na levi strani vsebuje enako število elementov kot terka na desni strani prireditvenega stavka. Kaj smo v zgornjem stavku pravzaprav naredili? Naredili smo terko spremenljivk, ki smo ji priredili terko na desni strani. Ker terki spremenljivk nismo dali nobenega imena, je v imenski prostor nismo shranili in zato tudi ni shranjena nikjer v pomnilniku. So pa v pomnilniku ostale spremenljivke, v katere smo razpakirali terko na desni. Kot smo videli že zgoraj pa lahko oklepaje pri definiciji tudi izpustimo. Torej lahko napišemo tudi nekaj takega

```
spremenljivka1, spremenljivka2, ... = terka
```

Mimogrede, tako razpakiranje elementov bi delovalo tudi, če bi imeli na desni strani prireditvenega stavka seznam.

Tak način razpakiranja elementov lahko uporabimo v našem zgledu z računanjem indeksa telesnih mas, s čimer se koda občutna skrajša:

```
for meritev in meritve:
    ime, visina, masa = meritev
    itm = masa/visina**2
    print("ITM osebe", ime, "je", itm)
```

Kodo lahko še dodatno skrajšamo, če razpakiranje naredimo kar v glavi zanke `for`:

```
for ime, visina, masa in meritve:
    itm = masa/visina**2
    print("ITM osebe", ime, "je", itm)
```

9.8 Pakiranje seznamov v sezname terk

Seznami terk, kot smo jih srečali zgoraj, torej predstavljajo lep način zapisovanja podatkov, ko želimo pri posamezni entiteti imeti več podatkov. Dejstvo pa je, da

velikokrat podatkov ne dobimo v taki obliki, ampak dobimo za vsako količino svoj seznam. Pri tem so sezname med seboj poravnani, kar pomeni, da istoležni elementi v vseh seznamih pripadajo isti entiteti. Elementi na indeksu 0 torej pripadajo entiteti 0, elementi na indeksu 1 entiteti 1 itd. V primeru imen, višin in mas, bi torej imeli tri seznime v obliki

```
imena = ["Janez", "Ana", "Nika"]
visine = [1.8, 1.65, 1.66]
mase = [75, 60, 55]
```

Elementi vseh treh seznamov torej na indeksu 0 pripadajo Janezu, na indeksu 1 Ani in na indeksu 2 Niki. Kaj imajo ti sezname skupnega? Indekse! Čez take podatke bi se torej lahko sprehodili tako, da se sprehajamo po indeksih in ne direktno po elementih. Naredimo torej sprehod z znako `for` od indeksa 0 do dolžine seznama - 1. Dolžine katerega seznama? Ni važno, saj so vsi enko dolgi (oziroma vsaj smiselno bi bilo, da so). To bi lahko naredili takole:

```
for i in range(len(imena)):
    ime = imena[i]
    visina = visine[i]
    masa = mase[i]
    itm = masa/visina**2
    print("ITM osebe", ime, "je", itm)
```

Kako pa bi lahko iz treh seznamov naredili seznam terk, s katerim smo delali zgoraj. Izkaže se, da se s takim problemom srečamo relativno pogosto, zato nam Python za *zapakiranje* več seznamov v seznam terk ponuja vgrajeno funkcijo `zip`. Funkcija `zip` iz zgornjih treh seznamov naredi točno to, kar bi si želeli:

```
>>> meritve = zip(imena, visine, mase)
>>> meritve
<zip object at 0x0000019A2A865D48>
```

Tale izpis je malo čuden, ampak ni z njim nič narobe. Funkcija `zip` je t.i. *iterator*, ki dejanski element seznama terk vrne, šele ko ga potrebujemo, oziroma posamezne elemente seznama vrača sproti. Če bi želeli imeti lepši izpis, bi lahko do njega prišli tako, da rezultat funkcije `zip` eksplicitno pretvorimo v seznam s funkcijo `list`:

```
>>> meritve = list(zip(imena, visine, mase))
>>> meritve
[('Janez', 1.8, 75), ('Ana', 1.65, 60), ('Nika', 1.66, 55)]
```

Sprehod čez seznime lahko torej naredimo na podoben način kot v prejšnjem poglavju, le da prej seznime zapakiramo v seznam terk:

```
for ime, visina, masa in zip(imena, visine, mase):
    itm = masa/visina**2
    print("ITM osebe", ime, "je", itm)
```

9.9 Zahteva po nespremenljivosti

V določenih primerih Python zahteva uporabo nespremenljivih podatkovnih tipov. Nespremenljive podatkovne tipe moramo uporabiti, kadar želimo podatke shranjevati v množico (`set`) in kadar želimo nek podatek uporabiti kot ključ (`key`) slovarja (`dict`). Če želimo v takem primeru uporabiti več elementov, moramo namesto po seznamu poseči po terki. V teh primerih je torej uporaba terk obvezna. Več o množicah in slovarjih bomo izvedeli prav kmalu.

Drug primer, v katerem bi nespremenljivost bila zaželena (ne pa obvezna), je, ko ne želimo, da funkcija spremeni vrednosti podanega argumenta. Kot smo videli lahko funkcije spreminjajo vrednosti svojih argumentov, tako da bodo spremembe vidne tudi izven funkcij. Če bi radi zagotovilo, da se podan argument izven funkcije zagotovo ne bo spremenil, namesto spremenljivega seznama enostavno uporabimo nespremenljivo terko. V tem kontekstu si pogledajmo spodnji zgled.

Zgled 32 *Kakšna je vrednost spremenljivk `seznam1` in `terka1` po izvedbi spodnje kode in kakšen bo izpis programa?*

```

1  def spremeni(a, b):
2      a += b
3
4  seznam1 = [1,2,3]
5  seznam2 = [4,5,6]
6  spremeni(seznam1, seznam2)
7  print(seznam1)
8
9  terka1 = (1,2,3)
10 terka2 = (4,5,6)
11 spremeni(terka1, terka2)
12 print(terka1)
```

Rešitev 32 *Kot smo videli že prej, se sprememba, ki smo jo nad plitvo kopijo spremenljivke `seznam1` naredili znotraj funkcije, odraža tudi izven funkcije, saj je seznam spremenljiv podatkovni tip. Ko torej spreminjamo njegovo plitvo kopijo, s tem spreminjamo vse spremenljivke, ki nanj kažejo.*

Kaj pa se zgodi, ko funkcijo pokličemo s terko. Najprej se ustvari plitva kopija terke v lokalnem imenskem prostoru funkcije (spremenljivka `a`). Ker je terka nespremenljiv podatkovni tip, je ne moremo spreminjati. Ob njenem spreminjanju se zato ustvari globoka kopija, torej nov podatek v pomnilniku, ki odraža narejeno spremembo. Na ta podatek pa kaže zgolj lokalna spremenljivka `a`. Ko se funkcija zaključi, njen lokalni imenski prostor skupaj z lokalno spremenljivko `a` izgine (tako kot tudi spremenjena terka – globoka kopija terke, s katero smo funkcijo poklicali). V sled temu sprememba, ki smo jo naredili znotraj funkcije, izven funkcije ni vidna.

Po izvedbi funkcije *spremeni* ima spremenljivka *seznam1* spremenjeno vrednost $[1,2,3,4,5,6]$, spremenljivka *terka1* pa ostane taka, kot je bila pred klicem funkcije $((1,2,3))$. Izpis programa je torej

```
[1,2,3,4,5,6] # spremenjena vrednost
(1,2,3) # nespremenjena vrednost
```

Funkcijam, ki kot argumente sprejemajo spremenljive podatkovne tipe, spremenjenih vhodnih argumentov ni potrebno vračati, saj se bodo spremembe odražale tudi izven funkcije. Funkcije, ki kot argumente sprejemajo nespremenljive podatkovne tipe, morajo spremenjene vhodne argumente eksplicitno vrniti, saj so spremenjene vrednosti sicer za vedno izgubljene. Oba načina si pogledjmo v spodnjih zgledih. Najprej si pogledjmo kako se lotiti pisanja in uporabe funkcije, ki sprejema nespremenljive podatke.

Zgled 33 Napiši funkcijo *dodaj_AT_niz*, ki sprejme dve nukleotidini zaporedji zapisani kot niza in v prvo zaporedje doda vse ponovitve baz *A* in *T* v enakem zaporedju kot nastopajo v drugem nizu. Funkcijo uporabi na zaporedjih *'ATCG'* in *'AATGGAATGG'*, tako da bo prvo zaporedje po njeni izvedbi spremenjeno.

Rešitev 33 Funkcija sprejema in spreminja podatke tipa *str*, ki je nespremenljiv podatkovni tip. Če bomo vhodne argumente spreminjali znotraj funkcije, se te spremembe izven funkcije ne bodo odražale, kar pomeni, da mora funkcija vračati spremenjen niz. Napišimo jo.

```
def dodaj_AT_niz(zaporedje1, zaporedje2):
    for baza in zaporedje2:
        if baza in 'AT':
            zaporedje1 += baza
    return zaporedje1
```

Na koncu torej vrnemo spremenjeno zaporedje1. Kljub temu, da smo znotraj funkcije to spremenljivko spreminjali, spremembe izven funkcije ne bodo vidne.

Klic funkcije moramo izvesti na tak način, da bo spremenila vrednost prve spremenljive, s katero funkcijo kličemo. Kako to doseči? Enostavno tako, da rezultat funkcije priredimo tej spremenljivki. Takole:

```
>>> zaporedje1 = 'ATCG'
>>> zaporedje2 = 'AATGGAATGG'
>>> zaporedje1 = dodaj_AT_niz(zaporedje1, zaporedje2)
```

Na tak način smo vrednost spremenljivke *zaporedje1* spremenili.

Poglejmo si še kakšne so razlike pri delu s spremenljivimi podatki.

Zgled 34 Napiši funkcijo `dodaj_AT_seznam`, ki sprejme dve nukleotidini zaporedji zapisani kot seznama enoznakovnih nizov (baz) in v prvo zaporedje doda vse ponovitve baz *A* in *T* v enakem zaporedju kot nastopajo v drugem nizu. Funkcijo uporabi na zaporedjih `['A', 'T', 'C', 'G']` in `'A', 'A', 'T', 'G', 'G', 'A', 'A', 'T', 'G', 'G'`, tako da bo prvo zaporedje po njeni izvedbi spremenjeno.

Rešitev 34 Navodilo naloge je praktično enako kot prej, le da tokrat namesto nespremenljivih podatkovnih tipov uporabljamo spremenljive. To pomeni, da ni potrebe po tem, da funkcija vrača spremenjen rezultat, saj se bodo spremembe odražale že preko podanega argumenta. Koda je torej sledeča:

```
def dodaj_AT_seznam(zaporedje1, zaporedje2):
    for baza in zaporedje2:
        if baza in 'AT':
            zaporedje1.append(baza)
```

Tokrat funkcija ne vrača ničesar uporabnega, zato njenega rezultata nima smisla ničemur prirejati. Vse kar potrebujemo je klic funkcije z ustreznimi argumenti:

```
>>> zaporedje1 = ['A', 'T', 'C', 'G']
>>> zaporedje2 = ['A', 'A', 'T', 'G', 'G', 'A', 'A', 'T', 'G', 'G']
>>> dodaj_AT_seznam(zaporedje1, zaporedje2)
```

Prepričajmo se, če je vrednost spremenljivke `zaporedje1` res spremenjena:

```
>>> zaporedje1
['A', 'T', 'C', 'G', 'A', 'A', 'T', 'A', 'A', 'T']
```


10 Slovarji

10.1 Zakaj slovarji?

Zdaj smo se že približje spoznali z različnimi sekvenčnimi podatkovnimi tipi, med katere smo uvrstili nize, sezname in terke. V spremenljivke, ki so pripadale tem tipom, smo lahko shranili več podatkov, pri čemer je bil podatek vedno povezan z določenim indeksom. Če se osredotočimo na sezname (kar bomo povedali velja sicer tudi za nize in terke), lahko rečemo, da so podatki v njih na nek način urejeni, ko smo v sezname dodajali nove podatke, smo te ponavadi dodajali na konec seznama, iskanje pa je potekalo tako, da smo se morali z zanko `for` sprehoditi čez celoten seznam in iskati dokler elementa nismo našli. Sezname smo lahko tudi sortirali po nekem ključu (recimo glede na relacijo $<$). Pri tem je bil rezultat sortiranja tak, da smo na manjših indeksih dobili elemente, ki so imeli manjšo vrednost od tistih na večjem indeksu, pri čemer je bila upoštevana uporabljena relacija.

V določenih primerih pa je bolj priročno, če lahko do vrednosti v spremenljivki dostopamo še preko česa drugega kot njenega indeksa. Pomislimo npr. na telefonski imenik števil (npr. v našem mobilnem telefonu), kjer lahko do telefonske številke neke osebe, dostopamo preko imena te osebe. Rekli bi lahko, da je ime osebe *ključ* preko katerega dostopamo do telefonske številke oziroma *vrednosti*. Na podoben način iščemo gesla v Slovarju slovenskega knjižnjega jezika (ali pač kakšnem drugem slovarju), kjer kot ključ nastopa iskana beseda oziroma geslo, kot vrednost pa razlaga iskane besede. Za razlago določene besede nam torej ni potrebno preiskati celotnega slovarja, ampak njeno razlago poiščemo preko gesla, ki ga je v slovarju načeloma enostavno (in predvsem hitro) najti.

V takih primerih lahko uporabimo v Python že vgrajen podatkovni tip `dict` (angl. *dictionary*) oziroma po slovensko kar *slovar*. V primeru slovarjev posamezna *vrednost* (angl. *value*) ni vezana na določen indeks, ampak je povezana na *ključ* (angl. *key*). Elemente slovarja torej vedno podajamo kot pare *ključ:vrednost*. Kakšna je prednost takega načina shranjevanja podatkov? Uporabno je takrat, ko ključ, po katerih shranjujemo in kasneje iščemo vrednosti, poznamo. V tem primeru je iskanje zelo hitro, saj nam ni potrebo preiskati celotnega slovarja, ampak slovarju zgolj podamo ključ in do vrednosti pridemo takoj. Operacija iskanja je torej zelo hitra v primerjavi s seznamami ali terkami.

10.2 Kako uporabljamo slovarje?

Slovarje zapisujemo z zavirami oklepaji, torej { in }, pri čemer elemente naštejemo kot pare ključ in vrednost ločene z dvopičjem (:). Prazen slovar bi naredili takole

```
>>> prazen_slovar = {}
```

Slovar, ki vsebuje enostaven telefonski imenik bi lahko zapisali kot

```
>>> imenik = {"Janez": "083455544",
               "Ana": "084566732",
               "Nika": "099563123"}
```

V tem primeru so ključi slovarja nizi, ki predstavljajo imena oseb, preko katerih lahko pridemo do vrednosti, ki v tem primeru predstavljajo telefonske številke zapisane v obliki nizov.

Mimogrede, slovar je spremenljiv podatkovni tip, kot smo zapisali že v tabeli 9.1.

10.3 Iskanje vrednosti

Telefonsko številko osebe lahko torej najdemo tako, da podamo njeno ime, kar je smiselno, saj imena svojih prijateljev ponavadi poznamo na pamet, njihovih telefonskih števil pa ne. Telefonsko številko od Janeza lahko npr. dobimo tako, da slovar indeksiramo po ključu "Janez":

```
>>> imenik["Janez"]
083455544
```

Indeksiranje se torej v primeru slovarjev namesto po indeksih izvaja po ključih. Kaj pa če bi želeli poiskati telefonsko številko od Dejana. Potem bi slovar indeksirali po ključu "Dejan", kar pa nam v konkretnem primeru vrne napako:

```
>>> imenik["Dejan"]
KeyError: 'Dejan'
```

Problem je v tem, da ključa "Dejan" v našem slovarju (še) ni, zato preko njega slovarja ne moremo indeksirati (tako kot nismo mogli indeksirati seznam z indeksom, ki ga v seznamu ni bilo). Obstoj ključa v slovarju lahko preverimo z operatorjem `in`:

```
>>> "Dejan" in imenik
False
>>> "Nika" in imenik
True
```

Preden do določenega ključa v slovarju dostopamo je torej smiselno, da preverimo, če ključ v slovarju sploh obstaja.

Zgled 35 *Napiši funkcijo `poisci`, ki kot argument sprejme imenik in ime osebe. Če imena ni v imeniku, naj funkcija vrne vrednost `False`, sicer pa njeno telefonsko številko.*

Rešitev 35 *Rešitev mora pred indeksiranjem po podanem ključu preveriti, če ključ v slovarju obstaja. V nasprotnem primeru vrne vrednost `False`.*

```
def poisci(imenik, ime):
    if ime in imenik:
        return imenik[ime]
    else:
        return False
```

10.4 Dodajanje in spreminjanje vrednosti

Indeksiranje po ključih, ki jih v slovarju ni, pa je v določenih primerih dovoljeno, in sicer takrat, ko želimo v slovar dodati nov par ključ, vrednost. Dodajanje novega elementa namreč izvedemo tako, da slovar indeksiramo po novem (neobstoječem) ključu in takemu indeksiranju priredimo vrednost, ki jo želimo s tem ključem povezati. Če bi npr. želeli v slovar dodati Dejana in z njim povezati neko telefonsko številko, npr. 089543678, bi to naredili takole:

```
>>> imenik["Dejan"] = "089543678"
```

V tem primeru napake nismo dobili, v slovarju pa se je ponavil nov par ključ, vrednost.

```
>>> imenik
{'Janez': '083455544', 'Ana': '084566732',
 'Nika': '099563123', 'Dejan': '089543678'}
```

Kaj pa se zgodi, če poskusimo v slovar dodati še enega Dejana? Ker v slovarju do vrednosti dostopamo preko ključev, morajo biti ključi enolični, kar pomeni, da se posamezen ključ lahko v slovarju pojavi največ enkrat. Če torej naredimo prireditev vrednosti preko ključa, ki v slovarju že obstaja, bomo s tem izvedli spreminjanje vrednosti, ki je povezana s tem ključem. Prireditev

```
>>> imenik["Dejan"] = "000000000"
```

bo torej spremenila telefonsko številko Dejana na 000000000.

```
>>> imenik
{'Janez': '083455544', 'Ana': '084566732',
 'Nika': '099563123', 'Dejan': '000000000'}
```

Če bi se želeli torej omejiti samo na dodajanje elementov, bi lahko prej preverili, če ključ preko katerega dodajamo v slovarju že obstaja in dodajanje izvedli le, če takega ključa še ni.

Zgled 36 *Napiši funkcijo **dodaj**, ki kot argumente prejme imenik, ime osebe in telefonsko številko osebe. Osebo in njeno telefonsko številko naj v slovar doda samo v primeru, če te osebe še ni v slovarju. Sicer naj izpiše, da ta oseba že obstaja.*

Rešitev 36 *V funkciji bomo tokrat izvedli prirejanje vrednosti, samo če podanega ključa še ni v slovarju. Poraja pa se dodatno vprašanje. Ali mora funkcija spremenjen slovar vračati? Odgovor je ne, saj je slovar spremenljiv podatkovni tip in spremembe slovarja, ki jih bomo v funkciji naredili in ki ga je funkcija prejela kot argument, se bodo odražale tudi izven funkcije.*

```
def dodaj(imenik, ime, stevilka):
    if ime not in imenik:
        imenik[ime] = stevilka
    else:
        print("Ta oseba že obstaja")
```

Če bi se želeli omejiti samo na spreminjanje elementov, bi morali spremeniti pogoj v stavku `if`, tako da številko prirejamo samo v primeru, če je ključ v slovarju že vsebovan.

Zgled 37 *Napiši funkcijo **spmeni**, ki kot argumente prejme imenik, ime osebe in telefonsko številko osebe. Telefonsko številko naj spremeni samo v primeru, če je oseba v imeniku že vsebovana. Sicer naj izpiše, da te osebe v imeniku ni.*

Rešitev 37 *Tokrat bomo v funkciji izvedli prirejanje vrednosti samo v primeru, če podan ključ v slovarju je. V nasprotnem primeru ne bi izvajali spreminjanja vrednosti za ključem, ampak bi izvajali dodajanje novega para ključ, vrednost.*

```
def spmeni(imenik, ime, stevilka):
    if ime in imenik:
        imenik[ime] = stevilka
    else:
        print("Te osebe ni v imeniku")
```

V določenih primerih, npr. pri štetju pojavitev nečesa, moramo v slovar dodati nov ključ, če tega ključa v slovarju še ni, ali spreminjati nanj vezano vrednost, če ta ključ v slovarju že obstaja. Poglejmo si sledeč primer:

Zgled 38 *Napiši funkcijo **prestej_baze**, ki kot argument prejme nukleotidno zaporedje baz zapisano kot niz. Funkcija naj vrne slovar, ki za posamezno bazo vsebuje število ponovitev.*

Rešitev 38 Za razliko od prej bo funkcija slovar vračala, saj ji ga kot argument nismo podali. V programu bi lahko predpostavljali, da delamo samo z bazami A, T, C in G. Tako bi si na začetku naredili slovar, v katerem kot ključi nastopajo oznake baz, nanje pa so vezane vrednosti 0. Takole:

```
baze = {'A':0, 'T':0, 'C':0, 'G':0}
```

Slabost takega pristopa je ta, da smo se omejili samo na ključe A, T, C in G. Kaj pa če kot vhod dobimo RNA zaporedje? Ali pa če se v zaporedju pojavi še kakšna oznaka, ki je nismo predvideli?

Boljši pristop bi bil, da na začetku naredimo prazen slovar in ko v nizu najdemo bazo, ki je v slovarju še ni, nanjo vežemo vrednost 1 (če smo jo našli v nizu prvič, potem to pomeni, da smo jo našli enkrat). V primeru, da v nizu najdemo bazo, ki v slovarju že obstaja, število njenih pojavitev povečamo za 1.

```
def prestej_baze(zaporedje):
    baze = {} # prazen slovar
    for baza in zaporedje:
        if baza not in baze: # baza v slovarju?
            baze[baza] = 1 # ne
        else:
            baze[baza] += 1 # da
    return baze
```

10.5 Brisanje vrednosti

V določenih primerih želimo elemente iz slovarjev tudi brisati. To lahko naredimo z besedico `del`, ki smo jo srečali že pri brisanju elementov iz seznama. Podobno kot pri seznamih tudi iz slovarjev brišemo z indeksiranjem vrednosti, ki jo želimo izbrisati:

```
del slovar[kljuc]
```

Tudi tokrat je smiselno, da pred brisanjem preverimo, če podan ključ v slovarju obstaja (brisanje po neobstoječem ključu bo spet vrnilo napako).

Zgled 39 Napiši funkcijo *izbrisi*, ki kot argumente prejme imenik in ime osebe, ki jo želimo iz imenika izbrisati. Brisanje naj se izvede samo v primeru, ko oseba v imeniku obstaja. Sicer naj funkcija izpiše, da te osebe v imeniku ni.

Rešitev 39 Spet bomo najprej preverili obstoj ključa, nato pa izvedli brisanje.

```
def izbrisi(imenik, ime):
    if ime in imenik:
        del imenik[ime]
```

```

else:
    print("Te osebe ni v imeniku")

```

10.6 Ključi in vrednosti

Posamezen ključ se lahko v slovarju pojavi največ enkrat. Ključi so torej enolični identifikatorji, preko katerih pridemo do posamezne vrednosti. Za ključe pa velja tudi to, da morajo biti nespremenljivi. Zakaj? Vrednost, ki je vezana na posamezen ključ, se zapiše na lokacijo v pomnilniku, ki je določena s preslikavo ključa v pomnilniško lokacijo (angl. *hash function*). Če ključ spreminjamo, se bo spremenila tudi vrednost preslikave. Za pravilno delovanje torej ključev ne smemo spreminjati (ko so enkrat shranjeni v slovarju). Zato je smiselno, da so ključi nespremenljivi podatki. Za vrednosti ni nobene omejitve – uporabimo lahko poljuben podatkovni tip vključno z drugim, ugnезdenim, slovarjem.

Do vseh ključev v slovarju lahko pridemo z uporabo metode `keys`, do vseh vrednosti z uporabo metode `values`, do vseh parov pa z uporabo metode `items`.

```

>>> imenik.keys()
dict_keys(['Janez', 'Ana', 'Nika', 'Dejan'])
>>> imenik.values()
dict_values(['083455544', '084566732', '099563123',
'089543678'])
>>> imenik.items()
dict_items([('Janez', '083455544'), ('Ana', '084566732'),
('Nika', '099563123'), ('Dejan', '089543678')])

```

Te metode vračajo nekaj kar je zelo podobno seznamom. Pri metodi `keys` tako dobimo seznam ključev, pri metodi `values` seznam vrednosti, pri metodi `items` pa seznam terk. Nad rezultatom, ki ga posamezna metoda vrne, se lahko sprehajamo z zanko `for`. Zanko `for` pa lahko izvedemo tudi direktno nad slovarjem, pri čemer se bomo tako sprehajali čez ključe slovarja:

```

>>> for k in imenik():
    print(k)

Janez
Ana
Nika
Dejan

```

Preko ključev seveda lahko dostopamo tudi do vrednosti.

Zgled 40 *Napiši funkcijo `izpisi_imenik`, ki kot argument prejme imenik. Funkcija naj izpiše vsebino imenika, tako da se vsak vnos nahaja v svoji vrstici, ime in telefonska številka pa naj bosta ločena z dvopičjem.*

Rešitev 40 V zanki `for` se lahko sprehajamo direktno čez slovar (sprehod čez ključe), čez ključe preko metode `keys` ali pa čez pare preko metode `items`. Uporabimo zadnjo možnost.

```
def izpisi_imenik(imenik):
    for ime, stevilka in slovar.items():
        print(ime, ":", stevilka)
```

Podoben sprehod bi lahko naredili tudi kadar npr. iščemo največjo vrednost v slovarju.

Zgled 41 Napiši funkcijo `naj_baza`, ki kot argument sprejme zaporedje baz, vrne pa ime baze, ki se v zaporedju pojavi največkrat. Pri tem si pomagaj s funkcijo `prestej_baze`.

Rešitev 41 Najprej bomo poklicali funkcijo `prestej_baze`, ki bo iz niza naredila slovar pojavitev baz. V naslednjem koraku moramo najti bazo, ki ima največ pojavitev. To bomo naredili na podoben način, kot smo izvedli iskanje največjega (ali najmanjšega) elementa v seznamu – s sprehodom z uporabo zanke `for`.

```
def naj_baza(zaporedje):
    baze = prestej_baze(zaporedje)

    naj_B = "" # naj baza
    M = 0 # število pojavitev
    for baza in baze:
        pojavitev = baze[baza]
        if pojavitev > M:
            naj_B = baza
            M = pojavitev
    return naj_B
```

10.7 Imenski prostor in slovarji

Poleg tega, da slovarje lahko navadni smrtniki uporabljamo za reševanje svojih programerskih problemov, slovarje za svoje delovanja uporablja tudi Python sam. Omenili smo že, da Python v svojem imenskem prostoru vsebuje vsa imena, preko katerih lahko dostopamo do vrednosti spremenljivk, funkcij in še česa drugega. Do imenskega prostora lahko pridemo s funkcijo `globals`. Poglejmo si, kaj ta funkcija vrne:

```
>>> globals()
{'__name__': '__main__', '__doc__': None, '__package__': None,
```

```
'__loader__': <class '_frozen_importlib.BuiltinImporter'>,
'__spec__': None, '__annotations__': {}, '__builtins__':
<module 'builtins' (built-in)>}
```

Funkcija vrne imena v imenskem prostoru in vrednosti, ki se za imeni skrivajo. Definirajmo novo spremenljivko in pokličimo gornjo funkcijo še enkrat.

```
>>> x = 1
>>> globals()
{'__name__': '__main__', '__doc__': None,
 '__package__': None,
 '__loader__': <class
 '_frozen_importlib.BuiltinImporter'>,
 '__spec__': None, '__annotations__': {},
 '__builtins__':
 <module 'builtins' (built-in)>, 'x': 1}
```

V imenskem prostoru se je zdaj očitno pojavilo ime `x`, za katerem se skriva vrednost 1. Če pogledamo bolj natančno, lahko vidimo, da je funkcija `globals` vrnila slovar.

```
>>> type(globals())
<class 'dict'>
```

V tem slovarju so ključi imena spremenljivk, funkcij itd., vrednosti pa tisto, kar se za imeni skriva. Ko v ukazno vrstico torej napišem `x` bo Python v svojem slovarju, ki predstavlja imenski prostor, pogledal, če tam obstaja ime `x` in vrnil vrednost, ki se za tem imenom skriva. To bi lahko naredili tudi takole:

```
>>> globals()['x']
1
```

Vsakič, ko definiram novo funkcijo ali spremenljivko, se njeno ime in vrednost doda v Pythonov slovar, podobno kot smo prej ključne in vrednosti v slovarje dodajali mi. Kaj se zgodi, ko določeni spremenljivki vrednost *povežemo* - vrednost za ključem z imenom te spremenljivke se enostavno poveže, saj lahko posamezen ključ v slovarju nastopa največ enkrat.

11 Množice

11.1 In še množice

Zadnji izmed vgrajenih podatkovnih tipov, ki si ga moramo pogledati, so množice. Množice za predstavitev podatkov uporabljamo takrat, ko želimo, da posamezen podatek obravnavamo največ enkrat. Ker nam Python poleg tega nad množicami omogoča izvedbo osnovnih operacij, kot so unija, presek in razlika, množice zelo spominjajo na matematične množice.

11.2 Uporaba množic

Množice podobno kot slovarje zapisujemo v zavite oklepaje, znotraj katerih naštejemo elemente. Množico elementov 1, 2 in 3, bi torej naredili takole:

```
>>> mnozica = {1,2,3}
>>> mnozica
{1, 2, 3}
>>> type(mnozica)
<class 'set'>
```

Kljub temu, da množice uporabljajo podoben zapis kot slovarji, ju Python med seboj brez problemov loči, saj slovarji za razliko od množic vsebujejo pare ključ: vrednost. Do težave pride le takrat, ko je množica oziroma slovar prazen. Do praznega slovarja pridemo tako, da podamo zavite oklepaje brez elementov:

```
>>> slovar = {}
>>> slovar
{}
>>> type(slovar)
<class 'dict'>
```

Do prazne množice pridemo z uporabo funkcije `set`, ki jo pokličemo brez argumentov:

```
>>> mnozica = set()
>>> mnozica
```

```
set()
>>> type(mnozica)
<class 'set'>
```

11.3 Omejitve pri uporabi množic

Elementi množic imajo zelo podobne omejitve kot ključi slovarjev, za katere prav tako velja, da lahko posamezno vrednost vsebujejo največ enkrat. Prav tako kot za ključe slovarjev tudi za množice velja, da lahko vsebujejo le nespremenljive podatkovne tipe.

Množice predstavljajo neurejeno strukturo, kar pomeni, da vrstni red elementov množice ni pomemben in tudi ni določen. Elementi torej niso vezani na indekse, zato množic ne moremo indeksirati in nad njimi delati rezine:

```
>>> mnozica = {1,2,3}
>>> mnozica[0]
TypeError: 'set' object is not subscriptable
```

Prav tako nad množicami ne moremo izvajati aritmetičnih operacij, kot smo jih npr. lahko izvajali nad nizi:

```
>>> {1,2,3}*3
TypeError: unsupported operand type(s) for *: 'set' and 'int'
>>> {1,2,3} + {4,5,6}
TypeError: unsupported operand type(s) for +: 'set' and 'set'
```

11.4 Osnovne operacije nad množicami

Kaj pa pravzaprav potem z množicami sploh lahko počnemo. Ko množico enkrat imamo se lahko čez njo sprehajamo z zanko `for`:

```
>>> mnozica = {1,2,3}
>>> for element in mnozica:
        print(element)
1
2
3
```

Poleg tega lahko preverjamo ali množica določen element vsebuje (ali ne) z operatorjema vsebovanosti:

```
>>> mnozica = {1,2,3}
>>> 1 in mnozica
True
```

```
>>> 2 not in mnozica
False
>>> 4 in mnozica
False
```

Številu elementov v množici pravimo tudi moč množice, do katere pridemo s klicem vgrajene funkcije `len`

```
>>> len({1,2,3})
3
```

Zgled 42 *Napiši funkcijo `razlicni`, ki kot argument sprejme niz, kot rezultat pa vrne število različnih znakov, ki v nizu nastopajo*

Rešitev 42 *Najprej bomo niz pretvorili v množico s funkcijo `set`. Ker množica posamezen element vsebuje največ enkrat, se bomo s tem znebili vseh potencialnih ponovitev znakov. Potem samo še izračunamo in vrnemo moč množice.*

```
1 def razlicni(niz):
2     mnozica = set(niz) # odstranitev duplikatov
3     return len(mnozica) # moč množice
```

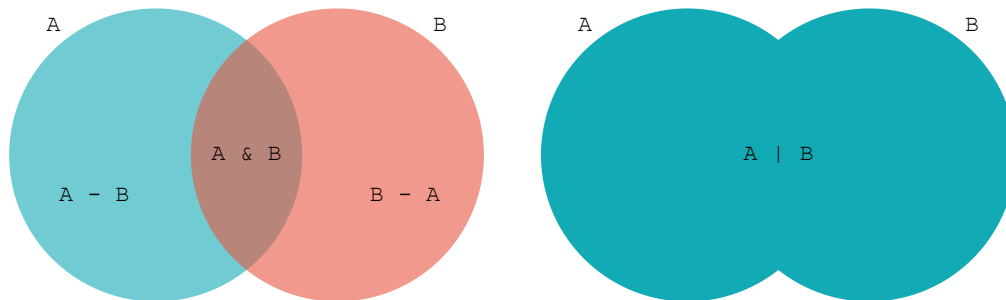
Tudi primerjalni operatorji so zdaj prilagojeni matematični interpretaciji množic. Ali je prva množica podmnožica druge, lahko npr. ugotovimo z operatorjem `<=`:

```
>>> {1,2,3} <= {1,2,3,4}
True
```

11.5 Presek, unija in razlika

Primeri najbolj tipičnih operacij, ki jih prikazujejo diagrami na sliki 11.1, so seveda presek ($A \& B$), unija ($A \mid B$) in razlika ($A - B$). Primer uporabe zgornjih operacij je sledeč:

```
>>> {1,2,3} & {3,4,5} # presek
{3}
>>> {1,2,3} | {3,4,5} # unija
{1,2,3,4,5}
>>> {1,2,3} - {3,4,5} # razlika
{1,2}
```



Slika 11.1 Vennova diagrama, ki ponazarjata osnovne operacije nad množicami. Slika na levi prikazuje razliko ($A-B$ in $B-A$) in presek ($A \& B$) med množicama A in B. Slika na desni prikazuje unijo ($A | B$) med množicama A in B

11.6 Metode množic: dodajanje in brisanje elementov

Dodajanje elementa v množico lahko izvedemo z uporabo metode `add`, ki kot argument sprejme element, ki ga želimo dodati:

```
>>> mnozica = {1,2,3}
>>> mnozica.add(4)
>>> mnozica
{1, 2, 3, 4}
```

V primeru, da dodajamo element, ki v množici že obstaja, metoda ne naredi ničesar. Dodajanje bi lahko izvedli tudi z operatorjem `|=`, ki naredi unijo množice z neko drugo množico. V tem primeru lahko dodamo več elementov naenkrat:

```
>>> mnozica = {1,2,3}
>>> mnozica |= {4,5}
>>> mnozica
{1, 2, 3, 4, 5}
```

Brisanje elementov iz množice vršimo z metodo `remove`. Ta v primeru neobstoja elementa vrne napako. Uporabimo lahko tudi metodo `discard`, ki v primeru neobstoja napake ne vrne, sicer pa deluje enako kot metode `remove`:

```
>>> mnozica = {1,2,3}
>>> mnozica.remove(3)
>>> mnozica
{1, 2}
>>> mnozica.remove(3)
KeyError: 3
>>> mnozica.discard(2)
>>> mnozica
{1}
```

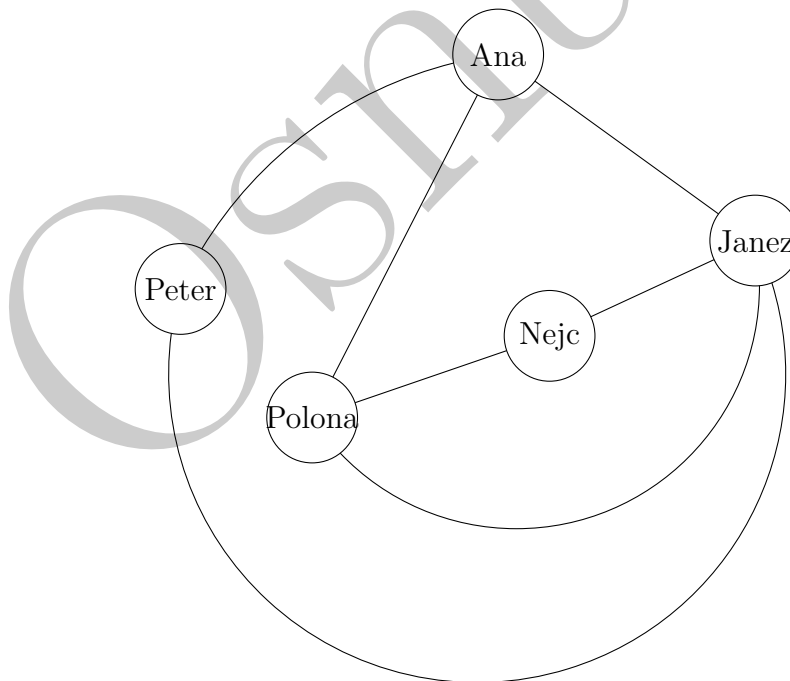
```
>>> mnozica.discard(2)
```

Podobno kot prej, lahko uporabimo tudi operator `-=`, ki naredi razliko med množico in neko drugo množico in to priredi izhodiščni množici:

```
>>> mnozica = {1,2,3}
>>> mnozica -= {2,3}
>>> mnozica
{1}
```

11.7 Zgled uporabe množic

Ker so bili sprotni zgledi v tem poglavju relativno skopi, bomo to nadoknadili z malo daljšim zgledom, s katerim bomo povadili tudi slovarje in mogoče še kaj. Zamislimo si, da bi radi opazovali omrežje prijateljev oziroma povezanost ljudi preko relacije *prijateljstvo*, bodisi v resničnem ali pa virtualnem življenju. Tako omrežje lahko predstavimo z *neusmerjenim grafom*, v katerem vozlišča predstavljajo imena oseb, povezave med vozlišči pa prijateljstva med osebami. Te povezave so neusmerjene, saj prijateljstvo deluje v obe smeri: če je oseba 1 prijatelj osebe 2, je tudi oseba 2 prijatelj osebe 1. Primer grafa prijateljstev prikazuje slika 11.2. V



Slika 11.2 Primer grafa omrežja prijateljstev.

grafu na sliki je npr. Ana prijateljica od Polone, saj sta med seboj povezani, Nejc pa ni prijatelj od Ane, saj med njima povezave ni.

Da se bomo lahko lotili nadaljnjih analiz omrežja prijateljstev, moramo tega najprej predstaviti v obliko, ki jo bomo lahko zapisali v Pythonu. V nadaljevanju bomo napisali par funkcij, ki jih bomo lahko uporabili pri posodabljanju omrežja in njegovi analizi.

Zgled 43 *Izberi si podatkovno strukturo, ki jo boš lahko uporabil pri predstavitvi in analizi omrežja prijateljstev ter z njo predstavi omrežje na sliki 11.2. Pri izbiri strukture upoštevaj to, da želiš nad omrežjem vršiti funkcije, kot so dodajanje prijateljev, brisanje prijateljev, iskanje skupnih prijateljev, iskanje osebe z največ prijatelji ipd.*

Rešitev 43 *Omrežje bi lahko predstavili na različne načine, a vendar težimo k tisti predstavitvi, ki nam bo pri nadaljnjih operacijah nad omrežjem prihranila največ dela, hkrati pa bodo operacije nad omrežjem potekale kar se da hitro. Začnimo z mogoče najbolj intuitivno, a ne preveč dobro predstavitvijo, in sicer s seznamom terk. V tem primeru vsaka terka v seznamu vsebuje ime osebe in seznam njenih prijateljev. Omrežje s slike 11.2 bi na tak način zapisali kot:*

```
prijatelji = [('Ana', ['Janez', 'Peter', 'Polona']),
              ('Janez', ['Ana', 'Nejc', 'Peter', 'Polona']),
              ('Nejc', ['Janez', 'Polona']),
              ('Peter', ['Ana', 'Janez']),
              ('Polona', ['Ana', 'Janez', 'Nejc'])]
```

Zakaj predstavitev ni najboljša? Zatakne se nam že pri izpisovanju vseh prijateljev podane osebe, saj že ta zahteva iskanje podane osebe s sprehodom čez (v najslabšem primeru) celoten seznam prijateljev. Če je ta seznam dolg, je lahko ta operacija časovno zelo potratna. Stvar bi lahko izboljšali, če bi namesto seznama terk uporabili kar slovar, v katerem so ključi imena oseb, vrednosti pa seznamami prijateljev. Takole:

```
prijatelji = {'Ana': ['Janez', 'Peter', 'Polona'],
              'Janez': ['Ana', 'Nejc', 'Peter', 'Polona'],
              'Nejc': ['Janez', 'Polona'],
              'Peter': ['Ana', 'Janez'],
              'Polona': ['Ana', 'Janez', 'Nejc']}
```

S tem smo rešili problem iskanja prijateljev podane osebe, saj do teh pridemo z enostavnim indeksiranjem po imenu osebe, ki nas zanima. Še vedno pa je problematično npr. iskanje skupnih prijateljev dveh oseb, saj zahteva ugnezdено zanko po seznamih prijateljev teh dveh oseb. Spet je ta operacija lahko časovno potratna, če so ti seznammi dolgi, kar v socialnih omrežjih zagotovo ni izključeno. Temu problemu bi se lahko izognili tako, da prijatelje podane osebe predstavimo z množicami, nad katerimi so operacije kot je presek bolj učinkovite. Poleg tega glede na podan opis

problema vrstni red prijateljev ni pomemben, posamezna oseba pa kot prijatelj druge osebe nastopa največ enkrat. Predstavitev bi bila torej sledeča:

```
prijatelji = {'Ana': {'Janez', 'Peter', 'Polona'},
              'Janez': {'Ana', 'Nejc', 'Peter', 'Polona'},
              'Nejc': {'Janez', 'Polona'},
              'Peter': {'Ana', 'Janez'},
              'Polona': {'Ana', 'Janez', 'Nejc'}}
```

Zgled 44 Napiši funkcijo `prijatelji_od`, ki kot argument sprejme omrežje prijateljev in ime osebe in vrne vse prijatelje podane osebe. V primeru, da podane osebe ni v omrežju, naj funkcija to izpiše.

Rešitev 44 Kot smo videli že prej, je zaradi izbrane predstavitve omrežja, iskanje prijateljev podane osebe enostavno. Preveriti moramo samo, če oseba v omrežju obstaja.

```
1 def prijatelji_od(prijatelji, oseba):
2     if oseba not in prijatelji: # ali je oseba v omrežju?
3         print("Ta oseba ne obstaja!")
4     else:
5         return prijatelji[oseba]
```

Zgled 45 Napiši funkcijo `dodaj_osebo`, ki kot argument sprejme omrežje prijateljev in ime osebe ter podano osebo doda v omrežje prijateljev s prazno množico prijateljev.

Rešitev 45 V omrežje bomo torej dodali novo osebo (nov ključ), na katerega bomo vezali prazno množico. Smiselno je tudi, da preverimo, če oseba v omrežju že obstaja. Ali mora funkcija vračati spremenjen slovar? Odgovor je seveda ne, saj je slovar spremenljiv podatkovni tip in se bo spreminjanje tega znotraj funkcije odražalo tudi izven funkcije (če bomo seveda slovar podali funkciji kot argument).

```
1 def dodaj_osebo(prijatelji, oseba):
2     if oseba not in prijatelji: # dodaj samo, če je še ni
3         prijatelji[oseba] = set() # prazna množica
```

Zgled 46 Napiši funkcijo `spoprijatelji`, ki kot argument sprejme omrežje prijateljev in imena dveh oseb, ki sta se spoprijateljili. Če sta osebi že v prijateljstvu, naj funkcija to izpiše. Če katerekoli izmed oseb še ni v omrežju, naj to osebo doda preko funkcije `dodaj_osebo`.

Rešitev 46 Spoprijateljjevanje bomo naredili tako, da bomo prvo osebo dodali v množico prijateljev druge osebe in obratno. Pri tem lahko uporabimo metodo `add`. Še prej bomo po potrebi posamezno osebo dodali v omrežje, če je tam še ni.

```

1 def spoprijatelji(prijatelji, oseba1, oseba2):
2     if oseba1 not in prijatelji:
3         dodaj_osebo(prijatelji, oseba1)
4     if oseba2 not in prijatelji:
5         dodaj_osebo(prijatelji, oseba2)
6
7     # spoprijatelji samo, če še nista prijatelja
8     if oseba2 not in prijatelji[oseba1]:
9         # dodaj oseba2 med prijatelje oseba1
10        prijatelji[oseba1].add(oseba2)
11        # dodaj oseba1 med prijatelje oseba2
12        prijatelji[oseba2].add(oseba1)
13    else:
14        print("Osebi sta že v prijateljstvu")

```

Primer klika:

```

>>> spoprijatelji(prijatelji, 'Ana', 'Nejc')
>>> prijatelji
{'Ana': {'Janez', 'Peter', 'Nejc', 'Polona'},
 'Janez': {'Peter', 'Ana', 'Nejc', 'Polona'},
 'Nejc': {'Janez', 'Ana', 'Polona'},
 'Peter': {'Janez', 'Ana', 'Polona'},
 'Polona': {'Janez', 'Peter', 'Ana', 'Nejc'}}
>>> spoprijatelji(prijatelji, 'Ana', 'Nejc')
Osebi sta že v prijateljstvu

```

Zgled 47 Napiši funkcijo *skregaj*, ki kot argument sprejme omrežje prijateljstev in imena dveh oseb, ki sta se skregali. Če osebi nista v prijateljstvu, naj funkcija to izpiše. Če katerekoli izmed oseb še ni v omrežju, naj funkcija to izpiše.

Rešitev 47 Pri skreganju bomo prvo osebo odstranili iz množice prijateljev druge osebe in obratno. Pri tem lahko uporabimo metodo *remove*. Še prej bomo preverili, če osebi sploh sta v omrežju in če sta v relaciji prijateljstva.

```

1 def skregaj(prijatelji, oseba1, oseba2):
2     if oseba1 not in prijatelji or oseba2 not in prijatelji:
3         print("Ene izmed oseb ni v omrežju!")
4         return # končaj in ne vrni nič
5     # ce oseba1 ni prijatelj oseba2, velja tudi obratno
6     if oseba1 not in prijatelji[oseba2]:
7         print("Osebi nista prijatelja!")
8         return # končaj in ne vrni nič
9

```



```

10     # odstrani prijateljstvo
11     prijatelji[oseba1].remove(oseba2)
12     prijatelji[oseba2].remove(oseba1)

```

Primer klica:

```

>>> skregaj(prijatelji, 'Ana', 'Nejc')
{'Ana': {'Janez', 'Peter', 'Polona'},
 'Janez': {'Peter', 'Ana', 'Nejc', 'Polona'},
 'Nejc': {'Janez', 'Polona'},
 'Peter': {'Janez', 'Ana', 'Polona'},
 'Polona': {'Janez', 'Peter', 'Ana', 'Nejc'}}
>>> skregaj(prijatelji, 'Ana', 'Nejc')
Osebi nista prijatelja!

```

Zdaj pa se lotimo še analize omrežja.

Zgled 48 Napiši funkcijo *najbolj_popularni*, ki kot argument sprejme omrežje prijateljstev in vrne ime osebe, ki ima največ prijateljev.

Rešitev 48 Spet rešujemo nalogo z iskanjem največjega elementa v slovarju, tokrat glede na dolžino množice, ki se skriva za posameznim ključem. Rešitev se ne bo dosti razlikovala od programa, ki je iskal oznako nukleotida, ki se v nukleotidnem zaporedju pojavi največkrat.

```

1 def najbolj_popularni(prijatelji):
2     naj_oseba = ""
3     naj_prijateljev = 0
4
5     for oseba in prijatelji: # sprehod čez ključe
6         # število prijateljev
7         prijateljev = len(prijatelji[oseba])
8
9         if prijateljev > naj_prijateljev:
10             naj_oseba = oseba
11             naj_prijateljev = prijateljev
12     return naj_oseba

```

Primer klica:

```

>>> najbolj_popularni(prijatelji)
'Janez'

```

Rešitev predpostavlja, da je enako najbolj popularna samo ena oseba, kar lahko rešimo z manjšo dopolnitvijo, tako da si najbolj popularne osebe shranimo v seznam.

```

1 def najbolj_popularni(prijatelji):
2     naj_osebe = []
3     naj_prijateljev = 0
4
5     for oseba in prijatelji: # sprehod čez ključe
6         # število prijateljev
7         prijateljev = len(prijatelji[oseba])
8
9         if prijateljev > naj_prijateljev:
10             naj_osebe = [oseba]
11             naj_prijateljev = prijateljev
12         elif prijateljev == naj_prijateljev:
13             naj_osebe.append(oseba)
14     return naj_osebe

```

Primer klika:

```

>>> spoprijatelji(prijatelji, 'Polona', 'Peter')
>>> najbolj_popularni(prijatelji)
['Janez', 'Polona']

```

Zgled 49 *Napiši funkcije `skupni_prijatelji`, `vsaj_od_enega` in `brez_njegovih`. Vse tri funkcije naj prejmejo omrežje prijateljstev in dve osebi ter vrnejo njune skupne prijatelje, prijatelje vsaj od ene izmed podanih oseb ter tiste, ki so prijatelji prve osebe ne pa prijatelji druge. V primeru, da v omrežju ni obeh podanih oseb, funkcije izpišejo, da vsaj ene podane osebe v omrežju ni.*

Rešitev 49 *V tej rešitvi bomo povadili uporabo preseka (`skupni_prijatelji`), unije (`vsaj od enega`) ter razlike (`brez njegovih`).*

```

1 def skupni(prijatelji, oseba1, oseba2):
2     if oseba1 not in prijatelji or oseba2 not in prijatelji:
3         print("Vsaj ene osebe ni v omrežju!")
4     else:
5         # presek
6         return prijatelji[oseba1] & prijatelji[oseba2]
7
8 def vsaj_od_enega(prijatelji, oseba1, oseba2):
9     if oseba1 not in prijatelji or oseba2 not in prijatelji:
10        print("Vsaj ene osebe ni v omrežju!")
11    else:
12        # unija
13        return prijatelji[oseba1] | prijatelji[oseba2]
14

```

```

15 def brez_njegovih(prijatelji, oseba1, oseba2):
16     if oseba1 not in prijatelji or oseba2 not in prijatelji:
17         print("Vsaj ene osebe ni v omrežju!")
18     else:
19         # razlika
20         return prijatelji[oseba1] - prijatelji[oseba2]

```

Primer klica:

```

>>> skupni(prijatelji, 'Ana', 'Janez')
{'Peter', 'Polona'}
>>> skupni(prijatelji, 'Ana', 'Jan')
Vsaj ene osebe ni v omrežju!
>>> vsaj_od_enega(prijatelji, 'Ana', 'Janez')
{'Janez', 'Peter', 'Ana', 'Nejc', 'Polona'}
>>> brez_njegovih(prijatelji, 'Ana', 'Janez')
{'Janez'}

```

Zgled 50 Napiši funkcijo `prijatelji_prijateljev`, ki kot argumenta sprejme omrežje prijateljev in ime osebe ter vrne prijatelje prijateljev podane osebe, pri čemer naj izpusti prijatelje podane osebe ter podano osebo samo.

Rešitev 50 Tokrat se bomo sprehodili čez prijatelje podane osebe z uporabo zanke `for` ter s prijatelji prijateljev dopolnili na začetku prazno množico. Potem bomo iz te množice izločili prijatelje podane osebe in osebo samo.

```

1 def prijatelji_prijateljev(prijatelji, oseba):
2     # prijatelji prijateljev je na začetku prazna
3     vsi_pp = set()
4
5     # sprehod čez prijatelje osebe
6     for prijatelj in prijatelji[oseba]:
7         pp = prijatelji[prijatelj]
8         # z unijo dodam prijatelje prijatelja
9         vsi_pp |= pp
10    # odstranim osebo
11    vsi_pp.remove(oseba)
12    # odstranim neposredne prijatelje osebe
13    vsi_pp -= prijatelji[oseba]
14
15    return vsi_pp

```

Primer klica:

```
>>> prijatelji_prijateljev(prijatelji, 'Nejc')  
{'Ana', 'Peter'}
```

Osnutek

12 Oblikovanje nizov

12.1 Delo z nizi

Nize smo do zdaj že dodobra spoznali. Preden se lotimo dela z datotekami pa si moramo pogledati še nekaj metod, ki jih lahko nad nizi uporabimo.

Ponovimo najprej osnovne operacije, ki smo jih do zdaj že izvajali nad nizi. Nad nizi lahko izvajamo zanko `for`, pri čemer v vsaki iteraciji zanke, dostopamo do enega znaka niza. Takole:

```
>>> niz = "abceda"
>>> for znak in niz:
    print(znak)
```

```
a
b
e
c
e
d
```

Nize lahko med sabo seštevamo, čemur smo rekli lepljenje ali konkatencija, poleg tega pa jih lahko množimo s celimi števili:

```
>>> "abc" + " " + "def"
'abc def'
>>> "abc" * 3
'abccabccabc'
```

Njihove elemente lahko tudi indeksiramo in delamo rezine. Ne moremo pa jih spreminjati. Odstranjevanje podniza z besedico `del` torej ne bo šlo skozi:

```
>>> niz = "abceda"
>>> niz[2:5]
'ece'
>>> del niz[2:5]
```

```
TypeError: 'str' object does not support item deletion
```

Zakaj ne? Ker so nizi nespremenljivi. Alternativa je seveda ta, da naredimo nov niz, ki odraža željeno spremembo. V zgornjem primeru bi torej lahko naredili nov

niz z lepljenjem dveh rezin. Eno pred podnizom, ki ga želimo odstraniti, in drugo za podnizom, ki ga želimo odstraniti. Takole:

```
>>> niz = "abceda"
>>> niz[:2] + niz[5:]
'abda'
```

Iz niza smo tako odstranili podniz "ece". Če želimo, da se sprememba odraža tudi preko spremenljivke z imenom `niz`, bomo nov (spremenjen) niz priredili temu imenu:

```
>>> niz = "abceda"
>>> niz = niz[:2] + niz[5:]
>>> niz
'abda'
```

Zapakirajmo vse skupaj v funkcijo.

Zgled 51 *Napiši funkcijo **odstrani**, ki sprejme dva niza in iz prvega niza odstrani prvo pojavitev drugega niza ter spremenjen niz vrne. Če se drugi podniz v prvem ne pojavi, naj funkcija vrne nespremenjen niz.*

Rešitev 51 *V funkciji moramo najprej preveriti, če in kje se v nizu nahaja podniz. To lahko naredimo tako, da se sprehajamo od začetka do konca niza in režemo rezine dolžine podniza. Da lahko delamo rezine, se moramo seveda sprehajati po indeksih niza. To bi izgledalo nekako takole:*

```
>>> niz = "abceda"
>>> podniz = "ece"
>>> for i in range(len(niz)):
        niz[i:i+len(podniz)]
'abe'
'bec'
'ece'
'ced'
'eda'
'da'
'a'
```

Znotraj zanke bomo preverjali, če je odrezana rezina enaka podnizu (zgoraj je to bilo v 3. iteraciji zanke). V tem primeru ga bomo odstranili (podobno kot prej) in vrnili rezultat. Napišimo celotno funkcijo.

```
1 def odstrani(niz, podniz):
2     for i in range(len(niz)):
3         rezina = niz[i:i+len(podniz)]
```

```

4         if rezina == podniz:
5             zacetek = i # začetek rezanja
6             konec = i + len(podniz) # konec rezanja
7             return niz[:zacetek] + niz[konec:]
8     return niz # nismo našli - vrni nespremenjen niz

```

Vprašanje je še ali mora funkcija res vračati spremenjen niz ali je dovolj, če niz, ki ga funkcija sprejme kot argument, spremenjamo znotraj funkcije (kot smo to delali pri seznamih, slovarjih in množicah). Spremenjen niz moramo seveda eksplicitno vrniti, saj je niz nespremenljiv. Spremembe, ki jih vršimo nad imenom, za katerim se skriva niz, se izven funkcije ne bodo ohranile, zato je potrebno nov (spremenjen) niz vrniti s stavkom *return*.

Tole je bilo še kar zakomplicirano. Izkaže pa se, da nam te in podobnih funkcij ni treba pisati, saj večinoma že obstajajo. Najdemo jih med metodami nizov. Izbrane (tj. tiste, ki jih bomo uporabljali pogosteje) si bomo podrobneje pogledali v nadaljevanju poglavja.

12.2 Iskanje podnizov

V prejšnjem zgledu se lahko uporabi *rezin* deloma izognemo z uporabo metode *startswith*, ki vrne vrednost *True*, če se niz, preko katerega smo metodo poklicali začne s podanim podnizom. Rešitev bi torej lahko nekoliko poenostavili:

```

1 def odstrani(niz, podniz):
2     for i in range(len(niz)):
3         if niz[i:].startswith(podniz):
4             zacetek = i # začetek rezanja
5             konec = i + len(podniz) # konec rezanja
6             return niz[:zacetek] + niz[konec:]
7     return niz # nismo našli - vrni nespremenjen niz

```

Poleg metode *startswith* lahko pri iskanju uporabimo še metodo *endswith*, ki preverja, če se niz konča s podanim podnizom, metodo *count*, ki prešteje število pojavitev podniza in metodo *find*, ki vrne lokacijo podniza oziroma *-1*, če podniza v nizu ni. Povadimo še uporabo metode *find*:

```

1 def odstrani(niz, podniz):
2     i = niz.find(podniz)
3     if i >= 0: # ali je podniz v nizu
4         zacetek = i # začetek rezanja
5         konec = i + len(podniz) # konec rezanja
6         return niz[:zacetek] + niz[konec:]
7     return niz

```

12.3 Odstranjevanje in spreminjanje (pod)nizov

Podnize pa lahko odstranimo zgolj s klicem metode `replace`, ki vse pojavitve prvega argumenta zamenja z drugim argumentom. Poskusimo

```
>>> niz = "abeceda"
>>> niz.replace("a","e")
'ebecede'
```

Metoda podanega niza ne spreminja. Seveda, saj ga ne more, ker so nizi nespremenljivi. Metoda torej niza ne more spremeniti, lahko pa vrne nov niz, ki odraža spremembo. Če želimo s klicem metode spremeniti vrednost, ki stoji za imenom spremenljivke, bomo rezultat klica priredili spremenljivki. Takole:

```
>>> niz = "abeceda"
>>> niz = niz.replace("a","e")
>>> niz
'ebecede'
```

Kako lahko uporabimo metodo pri odstranjevanju podnizov? Tako, da kot drugi argument podamo prazen niz.

```
>>> niz = "abeceda"
>>> niz.replace("a","")
'beced'
```

V navodilu prejšnje naloge je bilo zahtevno, da odstranimo samo prvo pojavitev podniza. Tudi to lahko rešimo z dodatnim argumentom, ki ga metoda `replace` sprejema. Poglejmo si izpis funkcije `help`:

```
>>> help("".replace)
Help on built-in function replace:
```

```
replace(old, new, count=-1, /) method of builtins.str instance
  Return a copy with all occurrences of substring old replaced
  by new.
```

```
    count
```

```
        Maximum number of occurrences to replace.
```

```
        -1 (the default value) means replace all occurrences.
```

```
    If the optional argument count is given, only the first count
    occurrences are replaced.
```

Podamo lahko tudi tretji argument `count`, ki pove koliko zamenjav naj metoda naredi (privzeto naredi vse zamenjave). Zdaj lahko našo funkcijo poenostavimo do konca:


```

1 def odstrani(niz, podniz):
2     return niz.replace(podniz, "", 1)

```

Pogosto uporabljeni metodi za spreminjanje nizov sta še metodi `upper` in `lower`, ki spremenita vse črke niza v velike oziroma male. Ker Python loči med velikimi in malimi črkami, lahko ti dve metodi uporabi, kadar te ločitve ne želimo imeti in vse pač spremenimo bodisi v velike bodisi v male črke. Povadimo na naslednjem zgledu.

Zgled 52 *Napiši funkcijo `palindrom`, ki kot argument sprejme niz in vrne `True`, če je niz palindrom in `False`, če ni. Pri tem naj funkcija ne loči med malimi in velikimi črkami.*

Rešitev 52 *Kako preveriti, če je niz palindrom, že znamo. Tokrat bomo dodali še to, da preverjanje ne loči med malimi in velikimi črkami. To bomo naredili tako, da bomo vse črke v nizu pred preverjanjem pretvorili v velike (ali pa v male).*

```

1 def palindrom(niz):
2     niz = niz.upper()
3     return niz == niz[::-1]

```

12.4 Razdruževanje in združevanje nizov

Funkcija `input` uporabnikov vnos vedno vrne zapisan kot niz. Prav tako bomo vsebino datotek (v naslednjem poglavju) vedno dobili najprej zapisano kot niz. V primeru, da uporabnik vnese eno število, lahko število zapisano kot niz enostavno pretvorimo preko funkcij `int` oziroma `float`. V primeru, da uporabnik naenkrat vnese več števil ali pa je vnos sestavljen iz števil in drugih znakov, ki jih mogoče želimo pretvoriti v kaj drugega pa ta pristop ni več mogoč. Deloma nas sicer lahko reši funkcija `eval`, katere uporaba zaradi varnosti v splošnem ni priporočljiva, poleg tega pa tudi ta slej ko prej odpove. V tem primeru moramo niz razčleniti (angl. *parse*) sami. Za ta namen lahko uporabimo metodo `split`, ki ji kot argument podamo ločilo (angl. *separator*), preko katerega naj niz loči v seznam nizov. Poglejmo si primer:

```

>>> niz = "jabolka,hruške,slive"
>>> niz.split(",")
['jabolka', 'hruške', 'slive']

```

Povadimo še na zgledu:

Zgled 53 *Napiši program, ki na podlagi uporabnikovega vnosa izračuna in izpiše vsoto podanih števil. Uporabnik bo števila vnesel v eni vrstici in jih med seboj ločil s podpičji.*

Rešitev 53 Program bo uporabnikov vnos (niz) ločil po podpičjih. Sledil bo sprehod čez dobljeni seznam, pretvorba nizov v števila in seštevanje.

```

1 vnos = input("Vnesi števila ločena s podpičji: ")
2 seznam = vnos.split(";")
3 s = 0 # začetna vsota
4 for st in seznam:
5     s += float(st) # pretvorba niza in prištevanje
6 print(s)

```

Zgled izvedbe programa:

```

Vnesi števila ločena s podpičji: 20;40.5;60
120.5

```

Metodo `split` lahko pokličemo tudi brez podanega ločila, pri čemer bo privzeto za ločilo uporabljen *prazen prostor* (angl. *white space*). Prazen prostor predstavlja znake, ki na zaslonu predstavljajo točno to – prazen prostor. Sem uvrstimo presledek, tabulator (`\t`) in novo vrstico (`\n`) (opomba: kombinacija znakov `\t` in `\n` pravzaprav predstavlja en sam znak). Dodatna prednost uporabe metode `split` brez argumenta je v tem, da bo več zaporednih ponovitev znakov za prazen prostor obravnaval kot eno samo. Poskusimo:

```

>>> niz = "1      2 \n 3\t4"
>>> niz.split(" ")
['1', '', '', '', '2', '\n', '3\t4']
>>> niz.split()
['1', '2', '3', '4']

```

Klic metode z argumentom praznega prostora očitno ni odstranil v celoti, klic metode brez argumentov iz niza uspešno pobral samo relevantne vrednosti.

Podobno, kot lahko podan niz glede na podano ločilo razbijemo na seznam nizov, lahko seznam nizov, glede na podano ločilo združimo v en niz. V tem primeru uporabimo metodo `join`, ki jo pokličemo nad nizom, ki predstavlja naše ločilo, kot argument pa ji podamo seznam nizov, ki ga želimo združiti. Takole:

```

>>> seznam = ["jabolka", "slive", "avokado"]
>>> ", ".join(seznam)
'jabolka, slive, avokado'

```

12.5 Odstranjevanje praznega prostora

Dopolnimo najprej naše iskanje palindromov.

Zgled 54 Napiši funkcijo *palindrom*, ki kot argument sprejme niz in vrne *True*, če je niz palindrom in *False*, če ni. Pri tem naj funkcija ne loči med malimi in velikimi črkami, poleg tega pa naj ignorira presledke, znake za nove vrstice in tabulatorje.

Rešitev 54 Želimo torej, da bi rešitev delovala tudi nad nečim takim kot je niz *"perica reže raci rep"*. Niz se prebere enako naprej kot nazaj, ampak samo v primeru, ko presledkov ne upoštevamo. Prva rešitev bo temeljila na tem, da vse bele prostore zamenjamo za prazne nize.

```
1 def palindrom(niz):
2     niz = niz.upper()
3     niz = niz.replace(" ", "")
4     niz = niz.replace("\n", "")
5     niz = niz.replace("\t", "")
6     return niz == niz[::-1]
```

Alternativa je, da najprej nad nizom uporabimo metodo *split* brez podanega argumenta, ki bo niz ločila v seznam nizov, pri čemer bo odstranila ves prazen prostor. Zatem uporabimo metodo *join*, pri čemer kot ločilo uporabimo prazen niz. Takole:

```
1 >>> niz = "perica \n reže \t raci rep"
2 >>> sez = niz.split()
3 >>> sez
4 ['perica', 'reže', 'raci', 'rep']
5 >>> niz = "".join(sez)
6 >>> niz
7 'pericarežeracirep'
```

Druga rešitev je torej sledeča:

```
1 def palindrom(niz):
2     niz = niz.upper()
3     sez = niz.split()
4     niz = "".join(sez)
5     return niz == niz[::-1]
```

Prve tri vrstice lahko združimo v eno:

```
1 def palindrom(niz):
2     niz = "".join(niz.upper().split())
3     return niz == niz[::-1]
```

Beli prostor lahko torej odstranjujemo na različne načine. V določenih primerih pa želimo beli prostor odstraniti samo pred začetkom in po koncu prave vsebine, vmes pa ne. V tem primeru lahko uporabimo metodo *strip*

```
>>> niz = "\n    \t Danes je lep dan!    \n"
>>> niz.strip()
'Danes je lep dan!'
```

12.6 Prilagajanje izpisa in formatiranje nizov

Funkcija `print` podane argumente združi v niz, ki ga izpiše na zaslon. Pri tem med argumente vstavi presledke (`sep =`), na koncu izpisa pa gre v novo vrstico (`end = "\n"`). Privzeto delovanje lahko spremenimo, tako da nastavimo (povozimo) privzete vrednosti izbirnima argumentom `sep` in `end`.

Včasih pa to ni dovolj in bi bili radi z izpisovanjem nekoliko bolj ustvarjalni. V tem primeru lahko uporabimo metodo `format`. Metodo `format` pokličemo nad nizom, ki vsebuje t.i. fiksne dele in dele, ki jih bo metoda zamenjala z argumenti, ki jih bomo podali ob klicu. Slednje znotraj niza označimo z zavitiimi oklepaji (`{` in `}`). Osnovni primer uporabe je sledeč:

```
>>> m = 1
>>> "{} meter/ov/a je {} centimeter/ov/a".format(m, m*100)
'1 meter/ov/a je 100 centimeter/ov/a'
```

Prvo pojavitev zavitega oklepaja je metoda `format` torej zamenjala s prvim argumentom, drugo pa z drugim. Znotraj zavutih oklepajev lahko tudi eksplicitno navedemo, kateri argument naj se uporabi pri zamenjavi. Takole:

```
>>> m = 1
>>> "{0} m = {1} cm, torej je {1} cm = {0} m".format(m, m*100)
'1 m = 100 cm, torej je 100 cm = 1 m'
```

Do malo gršega izpisa pride, kadar imamo veliko decimalk:

```
>>> cm = 12.673
>>> "{} cm = {} m".format(cm, cm/100)
'12.673 cm = 0.12673 m'
```

Rezultate lahko sicer zaokrožujemo z vgrajeno funkcijo `round`, ki ji kot argument podamo število decimalk. Alternativa je, da zaokroževanje podamo samo pri izpisu, tako da znotraj zavutih oklepajev malo bolj natančno povemo kako naj se izpis formatira. Takole:

```
>>> cm = 12.673
>>> "{:5.1f} cm = {:5.2f} m".format(cm, cm/100)
' 12.7 cm =  0.13 m'
```

Kaj smo s tem povedali? Z dvopičjem povemo, da želimo izpis malo oblikovati. Število 5 podaja (najmanjše) število mest, ki naj jih izpis zasede. Iz izpisa je vidno, da je metoda `format` pred posamezno število vstavila presledek, saj je izpisano

število dolgo 4 znake, mi pa smo povedali, da naj bo izpis dolg 5 znakov. Če bi bilo naše število daljše od 5 znakov, rezanja ne bi bilo, ampak bi `format` pač izpisal vse znake. Za piko smo podali število decimalk, ki naj bodo v izpisu. Pri centimetrih smo uporabili 1, pri metrih pa 2 decimalki. Oznaka `f` pomeni, da formatiramo decimalno število (angl. *float*). Oblikovanje celih števil in nizov je enostavnejše. V tem primeru podamo število mest, ki naj jih število (minimalno) zavzame.

```
>>> "{:3}: {:15}...{:5.2f}°C".format(1, "Ljubljana", 25.3)
' 1: Ljubljana          ...25.30°C'
```

Povemo lahko še ali naj se posamezen izpis poravna levo (<), desno (>) ali sredinsko (^):

```
>>> "{:<3}: {:<15}...{:>5.2f}°C".format(1, "Ljubljana", 25.3)
'1  : Ljubljana          ...25.30°C'
```

Presledke lahko zamenjamo tudi za kaj drugega, npr. za pike.

```
>>> "{:<3}: {:.<15}...{:>5.2f}°C".format(1, "Ljubljana", 25.3)
'1  : Ljubljana.....25.30°C'
```

Zdaj lahko uporabo formatiranja demonstriramo na celotnem zgledu:

```
>>> stevilke = [1,2,3]
>>> kraji = ["Ljubljana", "Maribor", "Nova Gorica"]
>>> temperature = [25.3, 21.32322, 26.433333]
>>> for st, kraj, temp in zip(stevilke, kraji, temperature):
    print("{:<3}: {:.<15}...{:>5.2f}°C".format(st, kraj, temp))
1  : Ljubljana.....25.30°C
2  : Maribor.....21.32°C
3  : Nova Gorica.....26.43°C
```

12.7 Formatiranje nizov in *f-Strings*

Novejša in hitrejša alternativa metodi `format` je formatiranje nizov zapisanih v obliki, ki ji rečemo *f-niz* oziroma *f-Strings*. Tovrstne nize zapisujemo tako, da pred začetkom niza (pred navednice) zapišemo črtko `f`. Python bo to razumel kot niz, ki ga mora še dodatno oblikovati. V osnovi tak niz zapišemo takole:

```
>>> f'niz'
'niz'
```

V *f-niz* lahko v zavite oklepaje vstavimo spremenljivke, ki jih bo Python pri izpisu zamenjal za njihove vrednosti, podobno kot pri metodi `format`:

```
>>> m = 1
>>> f"{m} meter/ov/a je {m*100} centimeter/ov/a"
'1 meter/ov/a je 100 centimeter/ov/a'
```

Tak način oblikovanja je nekoliko hitrejši kot oblikovanje z metodo `format`, predvsem pa je tak način oblikovanja bolj pregleden. Podobno kot pri metodi `format` lahko v zavite oklepaje podamo oblikovanje izpisa posamezne spremenljivke, le da je v tem primeru vrstni red malenkost drugačen, saj spremenljivko podamo kar v zavite oklepaje pred načinom njenega oblikovanja:

```
f"{spremenljivka:oblikovanje}"
```

Poskusimo na zgledu:

```
>>> cm = 12.673
>>> f"{cm:5.1f} cm = {cm/100:5.2f} m"
' 12.6 cm = 0.13 m'
```

Ostalo je zelo podobno oziroma enako kot pri uporabi metode `format`. Poskusimo še na zadnjem zgledu iz prejšnjega razdelka:

```
>>> stevilke = [1,2,3]
>>> kraji = ["Ljubljana", "Maribor", "Nova Gorica"]
>>> temperature = [25.3, 21.32322, 26.433333]
>>> for st, kraj, temp in zip(stevilke, kraji, temperature):
        print(f"{st:<3}: {kraj:..<15}...{temp:>5.2f}°C")
1  : Ljubljana.....25.30°C
2  : Maribor.....21.32°C
3  : Nova Gorica.....26.43°C
```

Rezultat je torej enak kot v primeru uporabo metode `format`, je pa koda postala nekoliko krajša in bolj pregledna.

13 Delo z datotekami

13.1 Zakaj pisati v datoteke in brati iz njih?

Programi, ki smo jih napisali do zdaj, so podatke hranili le za čas svojega izvajanja v t.i. *delovnem pomnilniku*. Ko se je nek program končal, so ti podatki izginili in ko smo program ponovno zagnali, smo morali te podatke ponovno zgenerirati, npr. tako, da smo jih prebrali od uporabnika (če smo bolj natančni, so podatki izginili, ko smo resetirali okolje IDLE npr. z zagonom drugega ali istega programa). Ko smo torej dopolnjevali telefonski imenik z novimi vnosi, se je imenik ob končanju programa izbrisal in dodani vnosi so izginili. To običajno predstavlja problem zaradi več razlogov:

- Ko se program neha izvajati, podatki izginejo. V primeru imenika se ta torej ob vsakem ponovnem zagonu programa resetira na začetno, npr. prazno, vsebino.
- Nimamo varnostnih kopij podatkov. Če se npr. telefon, ki uporablja naš imenik, ugasne, moramo celotno vsebino imenika ustvariti od začetka.
- Podatkov, ki smo jih ustvarili v enem programu, ne moremo oziroma težko uporabljamo v drugih programih.

Na srečo lahko podatke v svojih programih iz delovnega pomnilnika oziroma iz vsebine spremenljivk kadarkoli shranimo v t.i. *trajni pomnilnik* oziroma po domače na *disk*. S tem omogočimo njihovo trajno hrambo, kar pomeni, da lahko do teh podatkov pridemo tudi po zaključku izvajanja programa, na disku podatki ostanejo tudi po izklopu računalnika, do njih pa lahko pridemo tudi iz drugih programov. Kako pa podatke shranimo na disk? Podobno, kot smo do zdaj na disk shranjevali naše programe v obliki datotek s končnico `py`, lahko v (druge) datoteke shranjujemo tudi podatke, s katerimi delamo.

Tako shranjene datoteke lahko kasneje v svojih programih preberemo in s tem *obnovimo* stanje svojega delovnega pomnilnika oziroma določimo vrednosti spremenljivk na podlagi vsebine datoteke. Dodatna prednost branja podatkov iz datotek je to, da se nam zdaj ni treba več zanašati na vnos podatkov s strani uporabnika, ampak lahko podatke, ki jih želimo obdelati, v svoje programe preberemo kar iz datotek.

Če bi npr. želeli pregledati podatke o povprečnih mesečnih plačah v Sloveniji, najpogostejših imenih ali inflaciji, bi te lahko prenesli iz spletne strani Statističnega urada republike Slovenije (SURS, <https://www.stat.si/statweb>), jih v svojem programu prebrali in ustrezno obdelali. V tem poglavju si bomo pogledali kako podatke iz datoteke prebrati in kako v datoteko podatke zapisati.

13.2 Kaj je datoteka?

Datoteke predstavljajo osnovni način zapisovanja podatkov na disku (ali na kakšnem drugem mediju). Datoteke predstavljajo trajno hrambo podatkov, omogočajo prenašanje podatkov med različnimi uporabniki in lahko vsebujejo različne tipe podatkov od teksta, besedil, slik in filmov, do datotek, ki jih operacijskih sistem uporablja za svoje delovanje. Ponavadi tip datoteke označuje njena končnica. Programi v Pythonu na primer uporabljajo končnico `py`, slike pa končnice kot so `jpg`, `png` ali `gif`.

V grobem lahko datoteke ločimo v dve skupini, in sicer na *tekstovne datoteke* in *binarne datoteke*. V obeh primerih vsebino datoteke predstavlja zaporedje števil, ki si jih lahko v primeru tekstovnih datotek interpretiramo kot zaporedje znakov oziroma nek (neoblikovan) tekst. Primer tekstovne datoteke je recimo datoteka s končnico `py` ali datoteka s končnico `txt`. Za tekstovne datoteke velja, da jih lahko odpremo z beležnico ali beležnici podobnimi orodji (npr. *Notepad++*) in je njihova vsebina bolj ali manj berljiva. Kaj pa binarne datoteke? V skupino binarnih datotek uvrščamo vse ostalo, npr. slike, filme in oblikovana besedila. Če poskusite z beležnico odpreti datoteko s končnico `docx` (besedilo oblikovano z orodjem Microsoft Word), boste hitro videli, da to ni tekstovna datoteka. Tako kot pri tekstovnih datotekah je zapis s števili uporabljen tudi pri binarnih datotekah. Različna je le interpretacija teh števil. Medtem, ko lahko vsebino tekstovnih datotek orodja, kot je beležnica, enostavno dekodirajo kot zaporedje znakov, potrebujemo za binarne datoteke druga orodja oziroma programe, ki znajo njihovo vsebino dekodirati, glede na tip zapisa. Ena orodja pretvarjajo številke v slike, druga v filme, tretja v oblikovano besedilo. Katero orodje bo posamezno datoteko ob dvojnem kliku nanju odprlo, določa njena končnica. Če se končnica datoteke ne ujema z njeno vsebino, bo orodje vrnilo napako ali pa v najboljšem primeru prikazalo nekaj nenavadnega, podobno kot če datoteko s končnico `docx` odpremo z beležnico.

13.3 Tekstovne datoteke

Tekstovne datoteke torej vsebujejo neoblikovan tekst, ki lahko predstavlja na primer neko besedilo, program v Pythonu ali pa podatke, ki smo jih zgenerirali tekom delovanja našega programa. V okviru spoznavanja osnov programiranja se bomo neposredno ukvarjali zgolj s tekstovnimi datotekami.

13.4 Odpiranje datoteke

Če želimo neko datoteko v našem programu prebrati ali vanjo pisati, jo moramo najprej odpreti. To lahko naredimo z vgrajeno funkcijo `open`, ki ji bomo za začetek podali zgolj ime datoteke, do katere želimo dostopati. Naredimo tekstovno datoteko `stevila.txt` s sledečo vsebino¹:

```
1 4
2 5.6
3 2
4 100
5 15
```

Datoteko shranimo v mapo kamor pač najpogosteje shranjujemo svoje datoteke. Zdaj jo poskusimo odpreti v Pythonu:

```
>>> open('stevila.txt')
FileNotFoundError: [Errno 2] No such file or directory:
'stevila.txt'
```

Python pravi, da datoteke ne najde, čeprav smo jo pravkar ustvarili. Problem je v tem, da Python datoteko išče v svoji trenutni *delovni mapi*, ki očitno ni enaka tisti, kamor smo mi shranili svojo datoteko². Problem bi lahko rešili na dva načina.

Prvi način je, da Pythonu podamo *absolutno pot* do lokacije datoteke. Pythonu smo do datoteke v gornjem primeru podali t.i. *relativno pot*, kar pomeni, da bo datoteko iskal relativno glede na trenutno delovno mapo. Če je njegova delovna mapa npr. `C:\Windows\system32`, pričakuje, da se datoteka nahaja tu. Lahko pa Pythonu podamo celotno oziroma *absolutno pot*. Če smo datoteko npr. shranili v mapo `C:\Programiranje`, jo zdaj lahko odpremo takole:

```
>>> open('C:\\Programiranje\\stevila.txt')
<_io.TextIOWrapper
name='C:\\Programiranje\\stevila.txt' mode='r'
encoding='cp65001'>
```

Tokrat smo datoteko le našli. Mimogrede, zakaj je bilo pri podajanju poti potrebno podati dve poševnici (`\`). Kot smo videli v prejšnjem poglavju, poševnico Python obravnava kot začetek posebnega znaka. Z njo lahko npr. zapišemo tabulator (`\t`) ali pa znak za novo vrstico (`\n`). Če bi radi zapisali poševnico pa v niz enostavno vpišemo dve poševnici (`\\`).

Drugi način, ki je nekoliko enostavnejši in bolj pogost, je, da pač spremenimo svojo delovno mapo in potem do datoteke podamo relativno pot. To ponavadi pomeni,

¹Datoteko naredimo tako, da odpremo beležnico, vpišemo vsebino datoteke in datoteko shranimo.

²Do trenutne delovne mape lahko pridemo preko funkcije `getcwd` iz vgrajenega modula `os`.

da podamo samo ime datoteke (kot v primeru, ko je Python javil napako). Kako pa spremenimo trenutno delovno mapo? Načeloma nam za to ni potrebno posebej skrbeti, saj bo IDLE ob zagonu programa Pythonovo delovno mapo avtomatsko zamenjal za tisto, v kateri se program nahaja. Če bomo datoteko odprli iz programa, ki se bo nahajal v isti mapi kot datoteka, lahko torej podamo zgolj njeno ime. Če torej naredimo program `branje.py` in ga shranimo v mapo, v katero smo prej shranili `stevila.txt`, lahko iz njega datoteko odpremo takole:

```
open('stevila.txt')
```

13.5 Branje datoteke

Datoteko smo torej uspešno odprli, zdaj pa jo bomo poskusili prebrati. Dostopanje do datoteke lahko izvedemo preko objekta, ki ga vrne funkcija `read`. Če hočemo datoteko prebrati, moramo temu objektu prirediti neko ime, preko katerega ga bomo lahko kasneje še poklicali. Takole:

```
f = open('stevila.txt')
```

Do metod za delo z datoteko lahko zdaj dostopamo preko imena `f`. Prvi način za branje datoteke je uporaba metode `read`, ki prebere datoteko od začetka do konca in vrne niz z njeno vsebino. V našem primeru bo rezultat sledeč:

```
>>> f.read()
'4\n5.6\n2\n100\n15'
```

To je torej zapisano v datoteki. Niz vsebuje števila, ki so med seboj ločena z znaki za novo vrstico `\n`. Če bi želeli imeti lepši izpis, bi stvar izpisali s funkcijo `print`, ki znake za novo vrstico izpiše kot nove vrstice. Poskusimo:

```
>>> print(f.read())
```

Zakaj se tokrat ni nič izpisalo? Objekt za dostopanje do datoteke za imenom `f` beleži, do kje je bila datoteka že prebrana. Datoteko lahko namreč beremo tudi po kosih (npr. vrstico po vrstico), pri čemer nočemo, da bi vedno brali npr. samo prvo vrstico, ampak bi radi slej ko prej prišli do konca datoteke. Ker smo našo datoteko že prebrali, jo moramo za ponovno branje ponovno odpreti. Ker je trenutna delovna mapa zdaj že enaka mapi, kjer se nahaja naša datoteka, lahko odpiranje tokrat naredimo kar iz ukazne vrstice:

```
>>> f = open('stevila.txt')
>>> print(f.read())
4
5.6
2
100
15
```

Že prej smo omenili, da lahko datoteko beremo tudi po vrsticah. Lahko npr. uporabimo metodo `readline`:

```
>>> f.readline()
''
```

Tole spet ne deluje, ker se nahajamo na koncu datoteke. Ponovno jo odprimo in preberimo prvo vrstico:

```
>>> f = open('stevila.txt')
>>> f.readline()
'4\n'
```

Branje po vrsticah lahko izvedemo tudi z zanko `for`. Takole:

```
>>> for vrstica in f:
        print(vrstica)
5.6
2
100
15
```

Prva vrstica seveda manjka, ker smo jo prej prebrali že z metodo `readline`. Zakaj se v izpisu pojavlja prazna vrstica? En znak za prazno vrstico je prispevala funkcija `print`, drugega pa smo dobili iz prebrane datoteke. Znake za prazno vrstico lahko iz niza, ki predstavlja trenutno prebrano vrstico, odstranimo z metodo `strip`, ki odstrani prazen prostor pred začetkom in po koncu *prave* vsebine niza:

```
>>> f = open('stevila.txt')
>>> for vrstica in f:
        print(vrstica.strip())
4
5.6
2
100
15
```

Zadnja metoda, preko katere lahko preberemo datoteko, je metoda `readlines`, ki podobno kot metoda `read` prebere celo datoteko, jo razbije po vrsticah in te zapiše v seznam.

```
>>> f = open('stevila.txt')
>>> f.readlines()
['4\n', '5.6\n', '2\n', '100\n', '15']
```

Datoteko lahko torej preberemo na več različnih načinov, pri čemer je vsem skupno to, da vsebino datoteke preberejo kot niz. Če bi želeli vsebino obravnavati kot kaj drugega, jo bomo morali prej ustrezno obdelati (podobno kot moramo npr. v ustrezen podatkovni tip pretvoriti rezultat funkcije `input`).

Poskusimo branje na zgledu.

Zgled 55 *Napiši funkcijo `povprecje`, ki kot argument prejme niz z imenom datoteke in izračuna povprečje števil, ki so shranjena v datoteki. Predpostavljaš lahko, da je vsako število v datoteki zapisano v svoji vrstici. V primeru, da je datoteka prazna, naj funkcija vrne število 0.*

Rešitev 55 *Funkcijo lahko napišemo s sprehodom čez vrstice, pri čemer vsebino vsake vrstice prištevamo skupni vsoti, ki jo na koncu delimo s številom vrstic. Vsebino vrstice moramo prej seveda očistiti (metoda `strip`) in pretvoriti v število (funkcija `float`).*

```

1 def povprecje(ime_datoteke):
2     f = open(ime_datoteke)
3     s = 0
4     n = 0
5     for vrstica in f:
6         stevilo = float(vrstica.strip())
7         s += stevilo
8         n += 1
9     if n > 0:
10        return s/n
11    else:
12        return 0

```

Kaj pa če bi imeli v eni vrstici več števil? V tem primeru je potrebno vrstico prej še razbiti na posamezna števila. To lahko naredimo z uporabo metode `split`, ki jo pokličemo nad posamezno vrstico in ji kot argument podamo ustrezno ločilo.

Zgled 56 *Napiši funkcijo `povprecje`, ki kot argument prejme niz z imenom datoteke in izračuna povprečje števil, ki so shranjena v datoteki. Predpostavljaš lahko, da je v posamezni vrstici lahko zapisano več števil, ki so med seboj ločena z vejicami.*

Rešitev 56 *Tokrat bomo posamezno vrstico z metodo `split` razbili na seznam, čez katerega se bomo sprehodili z ugnezdjeno zanko `for`. Še vedno pa moramo posamezen element tega seznama pred prištevanjem skupni vsoti pretvoriti v število.*

```

1 def povprecje(ime_datoteke):
2     f = open(ime_datoteke)
3     s = 0

```

```

4     n = 0
5     for vrstica in f:
6         vrstica = vrstica.strip()
7         for element in vrstica.split(","):
8             stevilo = float(element)
9             s += stevilo
10            n += 1
11     if n > 0:
12         return s/n
13     else:
14         return 0

```

Poskusimo rešiti še težave, ki smo jih imeli s telefonskim imenikom (pa tega nismo vedeli).

Zgled 57 Napiši funkcijo `preberi_imenik`, ki kot argument prejme niz z imenom datoteke, ki vsebuje varnostno kopijo telefonskega imenika. Funkcija naj datoteko prebere in vrne slovar, v katerem so ključi imena oseb, vrednosti pa njihove telefonske številke zapisane kot nizi. Predpostavljaš lahko, da so podatki v datoteki zapisani tako, da posamezna vrstica vsebuje ime osebe, ki je od telefonske številke osebe ločeno z vejico. Pri tem se v eni vrstici nahajajo zgolj podatki ene osebe.

Rešitev 57 Podobno kot v prejšnjem zgledu bomo s posamezno vrstico naredili sledeče:

- odstranili bomo znak za novo vrstico z metodo `strip`;
- vrstico bomo razbili na seznam nizov glede na uporabljeno ločilo (',') z metodo `split`;
- seznam nizov bomo razpakirali in uporabili kot velevo navodilo naloge.

```

1 def preberi_imenik(ime_datoteke):
2     f = open(ime_datoteke)
3     imenik = {} # prazen slovar
4     for vrstica in f:
5         vrstica = vrstica.strip() # odstrani '\n'
6         seznam = vrstica.split(',')
7         ime = seznam[0]
8         stevilka = seznam[1]
9         imenik[ime] = stevilka # dodajanje vnosa
10    return imenik

```

*Funkcijo bi lahko še nekoliko izboljšali. Podobno, kot lahko razpakiramo terke, lahko razpakiramo tudi sezname. Torej lahko spremenljivkama **ime** in **stevilka** vrednosti priredimo kar v eni vrstici:*

```
ime, stevilka = vrstica.split(',')
```

Če želimo, da funkcija deluje, tudi v primeru, ko datoteka vsebuje prazne vrstice, moramo pred obdelavo posamezne vrstice preveriti, če ta slučajno ni prazna. Po odstranjevanju znaka za novo vrstico torej preverimo neenakost s praznim nizom:

```
if vrstica != '':
```

To lahko krajše zapišemo kar takole:

```
if vrstica:
```

Zapišimo celotno rešitev:

```
1 def preberi_imenik(ime_datoteke):
2     f = open(ime_datoteke)
3     imenik = {} # prazen slovar
4     for vrstica in f:
5         vrstica = vrstica.strip() # odstrani '\n'
6         if vrstica: # vrstica ni prazna?
7             ime, stevilka = vrstica.split(',') # razpakiraj
8             imenik[ime] = stevilka # dodajanje vnosa
9     return imenik
```

13.6 Pisanje v datoteko

Pisanje v datoteko lahko izvedemo z metodo **write**, ki ji kot argument podamo niz, ki ga želimo v datoteko zapisati. Poskusimo:

```
>>> f = open('stevila.txt')
>>> f.write(6)
TypeError: write() argument must be str, not int
```

Ravnokar smo rekli, da moramo metodi **write** podati **niz**, ki ga želimo zapisati. Če hočemo torej v datoteko zapisati število, ga moramo prej pretvoriti v niz s funkcijo **str**. Poskusimo še enkrat z nizom:

```
>>> f = open('stevila.txt')
>>> f.write('6')
io.UnsupportedOperation: not writable
```

Tokrat smo dobili drugo napako. Vidimo, da pisanje ni mogoče, ker smo datoteko odprli za branje. Več informacij o odprti datoteki, lahko izvemo preko spremenljivke, v katero smo datoteko odprli:

```
>>> f
<_io.TextIOWrapper name='stevila.txt' mode='r' encoding='cp65001'>
```

Iz izpisa vidimo, da smo odprli datoteko z imenom 'stevila.txt' (kar že vemo) in da je `mode` oziroma način odpiranja datoteke enak vrednosti 'r', ki se nanaša na branje oziroma *read*. Način odpiranja datoteke lahko funkciji `open` nastavimo z izbirnim argumentom `mode`, ki je privzeto enak vrednosti 'r'. Če argumenta ne nastavimo, bomo torej datoteko odprli za branje. Pogosto uporabljena načina odpiranja datoteke sta še 'w' oziroma pisanje (angl. *write*) in 'a' oziroma dodajanje (angl. *append*). Pisanju bi lahko, če bi bili nekoliko bolj natančni, rekli tudi prepisovanje, saj bomo s takim načinom odpiranja obstoječo datoteko prepisali. Dodajanju bi v podobnem kontekstu lahko rekli dopisovanje, saj bomo obstoječi datoteki vsebino dodajali (na koncu obstoječe vsebine). Pri obeh načinih pa zdaj napake, če datoteka s podanim imenom ne bo obstajala, ne bomo dobili, ampak bo Python to datoteko pač ustvaril. Poskusimo zdaj v datoteko 'stevila.txt' zapisati par števil. Najprej jo odpremo za pisanje:

```
>>> f = open('stevila.txt', mode = 'w')
```

Zdaj pa pišimo:

```
>>> f.write('1')
1
>>> f.write('2')
1
>>> f.write('3')
1
>>> f.write('45')
2
```

Vsakič, ko pokličemo metodo 'write' ta vrne število znakov, ki smo jih zapisovali. Ko zdaj datoteko `stevila.txt` odpremo z beležnico (ali beležnici podobnim orodjem), vidimo, da je ta prazna. Ko smo datoteko odprli za prepisovanje, je obstoječa vsebina izginila. Potem smo vanjo nekaj pisali. Zakaj je datoteka kljub pisanju še vedno prazna? Pravilo lepega dela z datotekami je, da jih po vsakem odpiranju tudi zapremo za kar lahko uporabimo metodo `close`. Pri branju to ni nujno potrebno, saj se bo datoteka zaprla avtomatsko, ko bo iz našega imenskega prostora izginila spremenljivka, preko katere smo datoteko najprej odprli in potem brali (najkasneje, ko bomo resetirali okolje IDLE). Pri pisanju pa se lahko zgodi, da bo vsebina datoteke brez eksplicitnega zapiranja datoteke ostala prazna. Ker je pisanje v datoteko počasna operacija, Python, kot smo že navajeni, to v določeni meri optimizira oziroma pohitri. Ko od njega zahtevamo pisanje v datoteko, namreč raje piše v medpomnilnik (angl. *buffer*), katerega vsebino bo v dejansko datoteko Python zapisal bodisi takrat, ko se bo nabralo malo več podatkov, da se mu bo splačalo to zapisati v datoteko, ali pa takrat, ko bomo datoteko zaprli, saj s tem

povemo, da vanjo ne bomo več pisali. Če želimo, da se naša števila v datoteko zagotovo zapišejo, jo moramo torej zapreti:

```
>>> f.close()
```

Ko zdaj odpremo datoteko, vidimo, da se je prejšnja vsebina datoteke prepisala – seveda, saj smo datoteko odprli za prepisovanje. V njej vidimo sledečo vsebino:

```
1 12345
```

To ni tisto kar smo hoteli? Mogoče ne, je pa to tisto, kar smo zahtevali. Metoda `write` namreč v datoteko zapiše točno tisto, kar od nje zahtevamo. Če bi želeli števila zapisati vsako v svojo vrstico, bi morali zraven vsakega števila zapisati tudi znak za novo vrstico (`\n`). Takole:

```
>>> f.write('1')
ValueError: I/O operation on closed file.
```

Ker smo datoteko zaprli, zdaj vanjo ne moremo več pisati. Pred ponovnim pisanjem, jo moramo torej spet odpreti:

```
>>> f = open('stevila.txt', mode='w')
>>> f.write('1\n')
2
>>> f.write('2\n')
2
>>> f.write('3\n')
2
>>> f.write('45\n')
3
```

Datoteko moramo na koncu seveda spet zapreti

```
>>> f.close()
```

Zdaj je njena vsebina sledeča:

```
1 1
2 2
3 3
4 45
5
```

Datoteka se konča s prazno vrstico, v katero smo skočili za nizom 45. Dodajmo v datoteko še par števil. Če želimo v datoteko novo vsebino dodajati k obstoječi vsebini, jo moramo odpreti za dopisovanje (`mode='a'`).

```
>>> f = open('stevila.txt', mode='a')
>>> f.write('55\n')
3
```



```
>>> f.write('133\n')
4
>>> f.write('-6.98\n')
6
>>> f.close()
```

Ko datoteko ponovno odpremo z beležnico, vidimo, da smo v njej dobili tri nove vrstice:

```
1 1
2 2
3 3
4 45
5 55
6 133
7 -6.98
8
```

Povadimo pisanje še na zgledu.

Zgled 58 *Napiši funkcijo **nakljucna**, ki kot argument prejme niz z imenom datoteke in celo število, ki podaja število naključnih števil, ki naj jih funkcija zapiše v datoteko s podanim imenom. Funkcija naj v vsako vrstico v datoteki zapiše največ 10 naključnih celih števil iz intervala od 1 do 100.*

Rešitev 58 *Naključna števila bomo generirali z uporabo funkcije **randint** modula **random**. V datoteko bomo pisali po eno število naenkrat. Po vsakem pisanju lahko zapišemo tudi presledek, na vsakih 10 pisanj pa bomo zapisali znak za novo vrstico.*

```
1 # uvažanje modulov vedno delamo izven funkcij
2 from random import randint
3
4 def nakljucna(ime_datoteke, n):
5     f = open(ime_datoteke, mode='w') # odpri za prepisovanje
6     for i in range(1,n+1): # zanko ponovimo n-krat
7         st = randint(1,100) # naključno število od 1 do 100
8         niz = str(st) # število moramo pretvoriti v niz
9         f.write(niz) # niz zapišemo v datoteko
10        if i % 10 == 0: # ali je čas za novo vrstico
11            f.write('\n')
12        else:
13            f.write(' ')
14    f.close() # datoteko na koncu zapremo
```

Metoda `write` torej v datoteko zapiše točno tisto, kar ji podamo, vedno pa ji moramo podati točno en podatek tipa `str`. Če bi želeli naenkrat zapisati več podatkov, jih moramo torej pred tem sami združiti. V določenih primerih bi bilo veliko bolj priročno, če bi lahko v datoteko pisali na enak način, kot smo stvari izpisovali na zaslon, tj. s funkcijo `print`. Ta sprejema vrednosti poljubnega tipa, poleg tega pa ji lahko podamo več vrednosti, ki jih bo med sabo združila (privzeto z ločilom `' '`). Izkaže se, da lahko za pisanje v datoteke uporabimo tudi funkcijo `print`. Pri tem moramo opcijskemu argumentu z imenom `file` prirediti spremenljivko, v katero smo datoteko odprli³ Poskusimo na prejšnjem zgledu.

Zgled 59 *Napiši funkcijo `nakljucna`, ki kot argument prejme niz z imenom datoteke in celo število, ki podaja število naključnih števil, ki naj jih funkcija zapiše v datoteko s podanim imenom. Funkcija naj v vsako vrstico v datoteki zapiše največ 10 naključnih celih števil iz intervala od 1 do 100. Za pisanje v datoteko uporabi funkcijo `print`.*

Rešitev 59 *Rešitev je zelo podobna kot prej, le da nam tokrat števil ni potrebno pretvarjati v nize, funkcija `print` pa tudi sama poskrbi za pisanje presledkov in novih vrstic (le pravilno jo moramo poklicati).*

```

1 # uvažanje vedno delamo izven funkcij
2 from random import randint
3
4 def nakljucna(ime_datoteke, n):
5     f = open(ime_datoteke, mode='w') # odpri za prepisovanje
6     for i in range(1,n+1): # zanko ponovimo n-krat
7         st = randint(1,100) # naključno število od 1 do 100
8         if i % 10 == 0: # ali je čas za novo vrstico
9             print(st, file=f) # nova vrstica
10        else:
11            print(st, end=' ', file=f) # presledek
12    f.close()
```

Za konec dopolnimo še zgled s telefonskim imenikom.

Zgled 60 *Napiši funkcijo `shrani_imenik`, ki kot argument prejme niz z imenom datoteke, ki naj vsebuje varnostno kopijo telefonskega imenika in slovar, v katerem so ključi imena oseb, vrednosti pa njihove telefonske številke. Funkcija zapiše imenik v datoteko s podanim imenom, in sicer tako, da vsaka vrstica vsebuje podatke o eni osebi, pri čemer je ime od telefonske številke ločeno z vejico.*

³Opcijski argument `file` funkcije `print` je privzeto postavljen na vrednost `sys.stdout`, kar pomeni, da pišemo na standardni izhod – v našem primeru je to konzola okolja IDLE.

Rešitev 60 V funkciji se bomo sprehodili čez cel imenik in v vsako vrstico zapisali ime in telefonsko številko. Ker sta ime in telefonska številka v slovarju shranjena kot niza, pretvorba v niz pred zapisovanjem ni potrebna.

```

1 def shrani_imenik(ime_datoteke, imenik):
2     f = open(ime_datoteke, mode='w') # prepisovanje
3     for ime, stevilka in imenik.items():
4         f.write(ime)
5         f.write(',')
6         f.write(stevilka)
7         f.write('\n')
8     f.close()

```

Štiri nize (*ime*, *,*, *stevilka* in *\n*) bi lahko združili skupaj in metodo *write* uporabili samo enkrat.

```

1 def shrani_imenik(ime_datoteke, imenik):
2     f = open(ime_datoteke, mode='w') # prepisovanje
3     for ime, stevilka in imenik.items():
4         f.write(ime+', '+stevilka+'\n')
5     f.close()

```

Za pisanje pa bi lahko uporabili tudi funkcijo *print*. Tej moramo spremeniti privzeto ločilo (presledek) v zahtevano ločilo (vejica). Takole:

```

1 def shrani_imenik(ime_datoteke, imenik):
2     f = open(ime_datoteke, mode='w') # prepisovanje
3     for ime, stevilka in imenik.items():
4         print(ime, stevilka, sep=',', file=f)
5     f.close()

```

13.7 Kodiranje znakov

Že na začetku poglavja smo omenili, da vsebino datoteke predstavlja zaporedje števil, ki jih lahko v primeru tekstovnih datotek pretvorimo v neoblikovan tekst oziroma neko zaporedje znakov. Vsak znak je torej shranjen kot neko število. *Kodiranje* določa preslikavo, ki posamezno število oziroma *kodo* preslika v določen znak. Do kode znaka lahko pridemo z uporabo vgrajene funkcije *ord*, ki ji kot argument podamo en znak:

```

>>> ord('a')
97
>>> ord('A')
65

```

Preslikavo lahko naredimo tudi v drugo smer (iz kode v znak), in sicer z uporabo funkcije `chr`, ki ji kot argument podamo kodo znaka:

```
>>> chr(97)
'a'
>>> chr(65)
'A'
```

Koda posameznega znaka je odvisna od uporabljenega kodiranja, pri čemer obstaja kar nekaj različnih načinov kodiranja. Izkaže se, da so osnovni znaki ne glede na uporabljeno kodiranje zakodirani na enak način (enaka koda predstavlja enak znak). Pri kodiranju teh se namreč uporablja 7-bitni standard ASCII (angl. *American Standard Code for Information Interchange*). Posamezen znak je torej predstavljen s 7 biti, kar pomeni, da lahko zakodiramo $2^7 = 128$ različnih znakov. Pri tem je 32 kod rezerviranih za posebne znake (med te sodi npr. tudi znak za novo vrstico). Torej nam ostane 96 kod, s katerimi moramo predstaviti male črke, velike črke, številke in različna ločila ter simbole. Izkaže se, da teh kod zelo hitro zmanjka in zato služijo zgolj predstavitvi osnovnih znakov in črk angleške abecede. Kaj pa naši šumniki, nemški preglasi in kitajske pismenke? Za predstavitev teh moramo uporabiti razširitve kodiranja ASCII. Omenimo dve razširitvi, ki sta v pogosti uporabi v današnjem času.

Kodne tabele (angl. *code page*) kodiranje ASCII razširjajo z dodatnim bitom, kar pomeni, da lahko z njihovo uporabo predstavimo dodatnih 128 znakov. To je sicer dovolj za znake č, ž in š, kaj pa za kitajske pismenke? Dodatnih kod seveda ni dovolj, da bi predstavili še te, poleg njih pa mogoče še črke iz cirilice, korejski hangul in še kaj drugega. Iz tega razloga se v različnih delih sveta uporabljajo različne kodne tabele. Pri nas je npr. v uporabi centralno-evropska kodna tabela cp1250, s katero lahko predstavimo tudi znake č, ž in š. Znak č je v tej kodni tabeli določen s kodo 232. V Nemčiji uporabljajo zahodno-evropsko kodno tabelo cp1252, s katero sicer ne moremo predstaviti šumnikov, lahko pa predstavimo preglase. V tej kodni tabeli koda 232 predstavlja znak è, kar ste mogoče že kdaj opazili, npr. pri predvajanju datoteke s podnapisi, ki jih program za predvajanje filmov poskuša dekodirati z napačnim kodiranjem.

Kljub temu, da se kodne tabele še vedno uporabljajo, je v današnjem času v večinski rabi standard *unicode*, katerega pomemben predstavnik je kodiranje UTF-8. Pri tem je posamezen znak zapisan z minimalno 8 biti. V primeru, da se v našem besedilu pojavijo znaki, ki ne nastopajo v osnovnem ASCII kodiranju (npr. č), se zapis posameznega znaka razširi (V konkretnem primeru na 16-bitni zapis). Prednost tega načina kodiranja je, da lahko datoteke iz različnih delov sveta, ki vsebujejo različne lokalne pisave, dekodiramo z enim samim kodiranjem in uporabniku ni potrebno ročno preklapljati med različnimi načini kodiranja, kot v primeru kodnih tabel.

Zakaj se moramo s kodiranjem znakov ukvarjati v okviru osnov programiranja? Pri

branju in pisanju datotek je zaželeno, da izrecno podamo tudi kodiranje, ki ga želimo uporabiti. S tem se namreč izognemo težavam z napačno interpretacijo znakov. Če je nekdo zapisal datoteko z uporabo kodiranja UTF-8, mi pa jo poskusimo odpreti s kodiranjem cp1250, bo prišlo do napake ali pa bodo znaki dekodirani narobe. Če kodiranja ne podamo (kot ga nismo podali v primerih zgoraj), bo uporabljeno privzeto kodiranje našega operacijskega sistema. Poglejmo si še enkrat kaj se skriva za spremenljivko, preko katere odpremo datoteko:

```
>>> f = open('stevila.txt')
>>> f
<_io.TextIOWrapper name='stevila.txt' mode='r' encoding='cp65001'>
```

Kodiranje (angl. *encoding*) je v konkretnem primeru nastavljeno na 'cp65001', ki predstavlja nekakšen ekvivalent kodiranja UTF-8 v operacijskem sistemu Windows. Operacijski sistem Windows sicer pogosto uporablja kodiranje cp1250, Linux pa UTF-8. Zmešnjava je torej popolna. To lahko povzroči nemalo težav, zato je pri delu z datotekami pomembno, da vedno podamo način kodiranja. Dogovorimo se, da bomo od zdaj naprej pri delu z datotekami vedno podali način kodiranja, ki naj bo kar UTF-8.

Kodiranje, ki ga želimo uporabiti pri pisanju ali branju datoteke, navedemo pri odpiranju datoteke s funkcijo `open` preko izbirnega argumenta `encoding`. Zapišimo datoteko z uporabo kodiranja UTF-8⁴:

```
>>> f = open("posebni_znaki.txt", mode='w', encoding="utf8")
>>> f.write("čžš")
3
>>> f.close()
```

Zdaj datoteko preberimo z uporabo kodiranja cp1250:

```
>>> f = open("posebni_znaki.txt", encoding="cp1250")
>>> f.read()
'ÄŤĹĹ'
```

Rezultat je seveda napačen. Poskusimo še s pravilnim kodiranjem:

```
>>> f = open("posebni_znaki.txt", encoding="utf8")
>>> f.read()
'čžš'
```

Izbira ustreznega kodiranja pri delu z datotekami je torej pomembna. Dopolnimo naši funkciji za branje telefonskega imenika iz datoteke in njegovo zapisovanje v datoteko, saj je velika verjetnost, da bodo v imenih naših prijateljev nastopali tudi šumniki.

⁴Argument `encoding` bomo nastavili na vrednost 'utf8', 'utf-8', 'UTF8' ali 'UTF-8'.

Zgled 61 *Napiši funkciji za branje in shranjevanje telefonskega imenika v in iz datoteke, pri čemer je imenik podan kot slovar, v katerem ključi predstavljajo imena oseb, vrednosti pa njihove telefonske številke. Datoteka naj bo zakodirana z uporabo kodiranja UTF-8.*

Rešitev 61 *Rešitev bo praktično enaka kot v prej, le da bomo pri odpiranju datoteke izbirni argument `encoding` postavili na vrednost `'utf8'`.*

```

1 def shrani_imenik(ime_datoteke, imenik):
2     # pri odpiranju datoteke podamo še kodiranje
3     f = open(ime_datoteke, mode='w', encoding='utf8')
4     for ime, stevilka in imenik.items():
5         print(ime, stevilka, sep=',', file=f)
6     f.close()
7
8 def preberi_imenik(ime_datoteke):
9     # pri odpiranju datoteke podamo še kodiranje
10    f = open(ime_datoteke, encoding='utf8')
11    imenik = {} # prazen slovar
12    for vrstica in f:
13        vrstica = vrstica.strip() # odstrani '\n'
14        if vrstica: # vrstica ni prazna?
15            ime, stevilka = vrstica.split(',') # razpakiraj
16            imenik[ime] = stevilka # dodajanje vnosa
17    return imenik

```

13.8 Datoteke CSV

Zapis CSV (angl. *comma separated values*) predstavlja način zapisovanja podatkov v tekstovne datoteke, pri čemer vsaka vrstica vsebuje svoj zapis (npr. osebo), podatki znotraj posamezne vrstice (npr. ime osebe in telefonska številka) pa so med seboj ločeni z vejicami. Primer datoteke CSV smo v tem poglavju že srečali, in sicer, ko smo v datoteko zapisovali telefonski imenik. Kot ločilo je lahko uporabljeno tudi kaj drugega kot vejica, npr. podpičje ali tabulator. Pomembno je, da je za ločilo uporabljen znak, ki se sicer med podatki nikoli ne pojavi. Če bi npr. datoteka vsebovala decimalna števila zapisana z decimalno piko, bi bila pika zelo slaba izbira za ločilo med podatki znotraj vrstice. Včasih uporabljeno ločilo nakazuje že končnica datoteke. Končnica `csv` npr. pogosto nakazuje na to, da je za ločilo uporabljena vejica (angl. *comma*) (čeprav to ni nujno), `tsv` pa tabulator. Včasih so datoteke CSV shranjene s končnico `txt`. V tem primeru je najlažje, če datoteko odpremo z beležnico ali beležnici podobnim orodjem in takoj vidimo, kateri znak je uporabljen za ločilo.

Prednost uporabe datotek CSV je v tem, da jih lahko zelo enostavno razčlenimo (angl. *parse*) oziroma z drugimi besedami, da lahko podatke iz njih zelo enostavno pridobimo in pretvorimo v obliko, ki bo najustreznejša za nadaljnjo obdelavo (spomnimo se npr. na omrežja prijateljstev). Kot smo videli v primeru branja telefonskega imenika, smo posamezno vrstico, ki predstavlja en zapis (npr. osebo) zgolj razbili z metodo `split`. Tako smo prišli do seznama podatkov (npr. imena in telefonske številke), ki pripadajo posameznemu zapisu (npr. osebi). Tudi zapisovanje podatkov datotek CSV je nadvse preprosto, saj moramo med podatke, ki jih zapisujemo zgolj vrniti izbrano ločilo, zapise pa med seboj ločiti z novimi vrsticami. Na zapise lahko torej gledamo tudi kot na vrstice, na podatke znotraj zapisov pa kot na stolpce v tabeli.

Datoteke CSV se v današnjem času uporabljajo zelo pogosto, saj zagotavljajo kompatibilnost zapisanih podatkov med različnimi orodji. Vse kar moramo za branje teh podatkov in izvedbo nadaljnjih analiz poznati je ločilo, ki loči podatke znotraj zapisa (tega lahko določimo tako, da datoteko odpremo z beležnico) in kodiranje, ki je bilo uporabljeno pri zapisovanju podatkov. V določenih primerih datoteka CSV vsebuje tudi *glavo*, ki podaja pomen posameznega stolpca v datoteki, v določenih primerih pa moramo vnaprej vedeti kaj posamezen stolpcev predstavlja. Datoteke CSV lahko enostavno uvozimo v različna orodja za obdelavo podatkov, kot je npr. Excel, lahko pa obdelavo podatkov naredimo kar v jeziku Python. Za shranjevanje in branje datotek CSV obstaja kar nekaj knjižnic (v zadnjem času se za namen obdelave podatkov, ki zajema tudi branje in pisanje datotek CSV uporablja predvsem knjižnica `pandas`, ki jo bomo še srečali), mi pa bomo branje malo kompleksnejše datoteke CSV izvedli kar z osnovnimi konstrukti jezika Python, podobno kot smo to naredili že s telefonskim imenikom. Poglejmo si naslednji primer.

Najprej bomo pridobili testne podatke, v obliki datoteke CSV.

Zgled 62 Na spletni strani Statističnega urada Republike Slovenije (<https://www.stat.si/>) poišči bazo podatkov SiStat, v njej pa podatke o povprečnih mesečnih bruto in neto plačah pri pravnih osebah javnega in zasebnega sektorja. Te izvozi v obliki CSV, pri čemer želimo imeti ločene podatke za javni in zasebni sektor ter za bruto in neto plače, zanimajo pa nas plače za posamezen mesec.

Rešitev 62 Na začetni strani poiščemo povezavo na podatkovno bazo SiStat (<https://pxweb.stat.si/SiStat>). Sledimo povezavam: Demografsko in socialno področje → Trg dela → Plače in stroški dela → Povprečne mesečne plače → Povprečne mesečne bruto in neto plače pri pravnih osebah javnega in zasebnega sektorja, Slovenija, mesečno. Tu si izberemo mesece, ki jih želimo izvoziti, sektorja in bruto ter neto plače. Pod Meritve si izberemo še Plače za mesec. Spodaj si izberemo eno izmed oblik CSV za izvoz in podatke izvozimo. Datoteko preimenujemo v `place.csv`. Primer tako izvožene datoteke je na voljo na povezavi.

Zdaj bomo poskusili podatke uvoziti v Python in izračunali povprečje razpoložljivih mesečnih plač v javnem in v zasebnem sektorju.

Zgled 63 *Napiši funkcijo `uvozi_place`, za uvoz podatkov o mesečnih plačah. Funkcija naj kot argument sprejme ime datoteke s podatki in slovar, ki kot ključ vsebuje naziv sektorja, kot vrednost pa nov slovar. V tem slovarju naj bodo ključni nizi `'mesec'`, `'bruto'` in `'neto'`, za katerimi so sezname, ki predstavljajo mesece, bruto zneske in neto zneske. Seznami naj imajo na enakem indeksu podatek za isti mesec.*

Rešitev 63 *Najprej moramo podrobneje pogledati kako je datoteka strukturirana. V ta namen jo odpremo z beležnico ali podobnim orodjem (ne z orodjem Excel). Če jo odpremo z orodjem Notepad++, lahko enostavno preverimo kako je datoteka zakodirana. Če določeni znaki niso berljivi, postanejo berljivi, ko kodiranje nastavimo na cp1250 (Encoding → Character Sets → Central European → Windows-1250). To kodiranje bomo torej uporabili pri odpiranju datoteke:*

```
f = open("place.csv", encoding="cp1250")
```

Zdaj se lotimo vsebine datoteke. Takoj vidimo, da prve tri vrstice ne vsebujejo podatkov. Te bomo torej preskočili. Lahko kar tako, da jih preberemo:

```
f.readline()
f.readline()
f.readline()
```

Ostale vrstice bomo shranjevali v slovar seznamov, ki ga bomo lahko kasneje uporabili za analizo mesečnih plač. Najprej si pripravimo prazen slovar:

```
place = {}
```

Zdaj se bomo z zanko `for` sprehodili čez preostanek datoteke. Posamezno vrstico bomo najprej oklestili znaka za novo vrstico, potem pa razbili glede na uporabljeno ločilo. V beležnici ali orodju Notepad++ vidimo, da je to tabulator (`\t`).

```
seznam = vrstica.strip().split("\t")
```

Tako pridobljen seznam bo sestavljen iz štirih podatkov: podatka o mesecu, sektorju, bruto plači in neto plači. Razpakirajmo ga:

```
mesec = seznam[0]
sektor = seznam[1]
bruto = seznam[2]
neto = seznam[3]
```

Šlo bi tudi hitreje:

```
mesec, sektor, bruto, neto = seznam
```


Podatku o mesecu in sektorju lahko pred nadaljnjo obdelavo odstranimo dvojne navednice:

```
mesec = mesec.replace('\"', '')
sektor = sektor.replace('\"', '')
```

Zneske bomo pretvorili v števila tipa *float*:

```
bruto = float(bruto)
neto = float(neto)
```

Potem bomo preverili, če je sektor že med ključi slovarja *place*. Če ga še ni, ga bomo dodali in nanj vezali začetni slovar za shranjevanje podatkov *'mesec': [], 'neto': [], 'bruto': []*. Takole.

```
if sektor not in place: # če sektorja še ni
    place[sektor] = {'mesec': [],
                    'neto': [],
                    'bruto': []}
```

Zdaj v seznam, ki pripadajo trenutnemu sektorju dodamo trenutne podatke o mesecu, bruto in neto plači:

```
place[sektor]['mesec'].append(mesec)
place[sektor]['neto'].append(neto)
place[sektor]['bruto'].append(bruto)
```

Celotno rešitev zapisano v obliki funkcije predstavlja spodnja koda:

```
1 def uvozi_place(ime_datoteke):
2
3     # datoteka je zakodirana s cp1250
4     f = open(ime_datoteke, encoding='cp1250')
5
6     # prve tri vrstice vržemo proč
7     f.readline()
8     f.readline()
9     f.readline()
10
11     # prazen slovar
12     place = {}
13
14     for vrstica in f:
15         # razbijmo vrstico na seznam s podatki
16         seznam = vrstica.strip().split("\t")
17         # razpakirajmo seznam
18         mesec, sektor, bruto, neto = seznam
```

```

19
20     # odstranimo dvojne navednice
21     mesec = mesec.replace('"', '')
22     sektor = sektor.replace('"', '')
23
24     # zneske pretvorimo v števila
25     bruto = float(bruto)
26     neto = float(neto)
27
28     if sektor not in place: # če sektorja še ni
29         place[sektor] = {'mesec': [],
30                           'neto': [],
31                           'bruto': []}
32
33     # dodamo podatke v ustrezen seznam
34     place[sektor]['mesec'].append(mesec)
35     place[sektor]['neto'].append(neto)
36     place[sektor]['bruto'].append(bruto)
37
38     # podatki so zdaj pripravljeni
39     return place

```

Podatke lahko zdaj preberemo in obdelamo. Lahko npr. izračunamo povprečje mesečnih bruto plač v javnem in v zasebnem sektorju:

```

>>> place = uvozi_place('place.csv')
>>> ZJ = place['Javni sektor']['neto']
>>> ZZ = place['Zasebni sektor']['neto']
>>> avg_javni = sum(ZJ)/len(ZJ)
>>> avg_zasebni = sum(ZZ)/len(ZZ)
>>> avg_javni
1886.8506756756758
>>> avg_zasebni
1502.0971621621622

```

Vidimo torej, da so plače v javnem sektorju precej višje. Z analizami bi lahko nadaljevali. Lahko bi npr. poiskali mesec, ko je bila plača najvišja, pogledali kakšni so trendi itd. Pred podrobnejšo obdelavo pa vselej pomaga, če si podatke prej narišemo. In ravno to bomo naredili v naslednjem poglavju.

14 Vizualizacija podatkov

14.1 Knjižnica Matplotlib in njena namestitve

Za vizualizacijo podatkov bomo uporabljali knjižnico Matplotlib, ki predstavlja osnovo za risanje kakršnihkoli grafov. Ker v osnovni različici jezika Python še ni nameščen, ga moramo pred uporabo namestiti (če imate nameščeno distribucijo Anaconda, imate to knjižnico že nameščeno). Kot smo spoznali v poglavju 8 lahko za namestitev knjižnice uporabimo orodje `pip`, tako da zaženemo ukazno vrstico svojega operacijskega sistema (ne okolja IDLE) in vanjo vpišemo

```
> pip install matplotlib
```

Podrobnejša navodila za namestitev paketov smo podali že v poglavju 8, zato jih tu ne bomo podvajali. Če ima vaš računalnik povezavo z internetom, bo orodje `pip` samo preneslo potrebne namestitvene datoteke in knjižnico namestilo.

Za vizualizacijo naših podatkov bomo uporabili Matplotlibov vmesnik (angl. *interface*) `pyplot`, ki nam risanje precej olajša. V svoje programe ga bomo uvozili takole:

```
import matplotlib.pyplot as plt
```

Zdaj lahko do funkcij za risanje grafov dostopamo takole:

```
plt.ime_funkcije(argumenti)
```

14.2 Funkciji `plot` in `show`

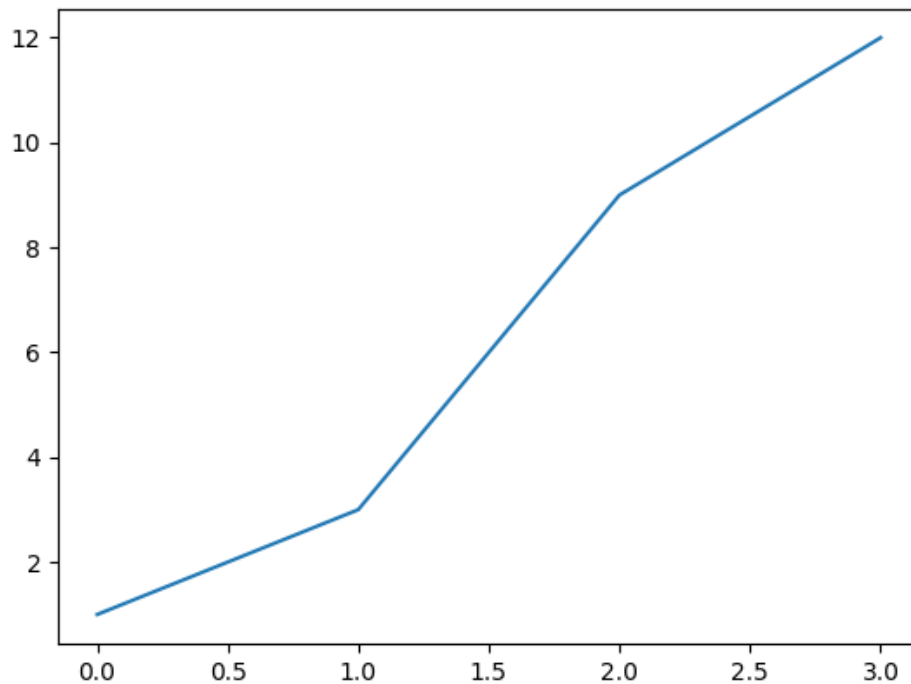
Začeli bomo s funkcijo `plot`, ki omogoča izris črtnega grafa (angl. *line plot*). V osnovi ji lahko podamo zgolj en seznam. Poskusimo:

```
>>> Y = [1, 3, 9, 12]
>>> plt.plot(Y)
[<matplotlib.lines.Line2D object at 0x000001DD7FB28860>]
```

Nekaj se je očitno zgodilo, grafa pa še vedno ne vidimo. Funkcija `plot` deluje tako, da grafe riše v ozadju in te dodaja na risalno površino, ki pa jo pokaže, šele ko pokličemo funkcijo `show`.

```
>>> plt.show()
```

Zdaj se je prikazal graf, ki ga prikazuje slika 14.1. Točke, ki smo jih podali



Slika 14.1 Črtni graf, pri čemer smo podali koordinate točk na osi y .

funkciji `plot`, očitno predstavljajo koordinate y narisane grafa. Točke na osi x je Matplotlib določil kar glede na indekse točke v seznamu `Y`. To lahko spremenimo, tako da funkcijo `plot` pokličemo z dvema seznamoma, pri čemer prvi določa koordinate točk na osi x drugi pa koordinate točk na osi y ¹. Seznama morata biti seveda enakih dolžin. Do enakega grafa kot zgoraj, bi lahko prišli takole:

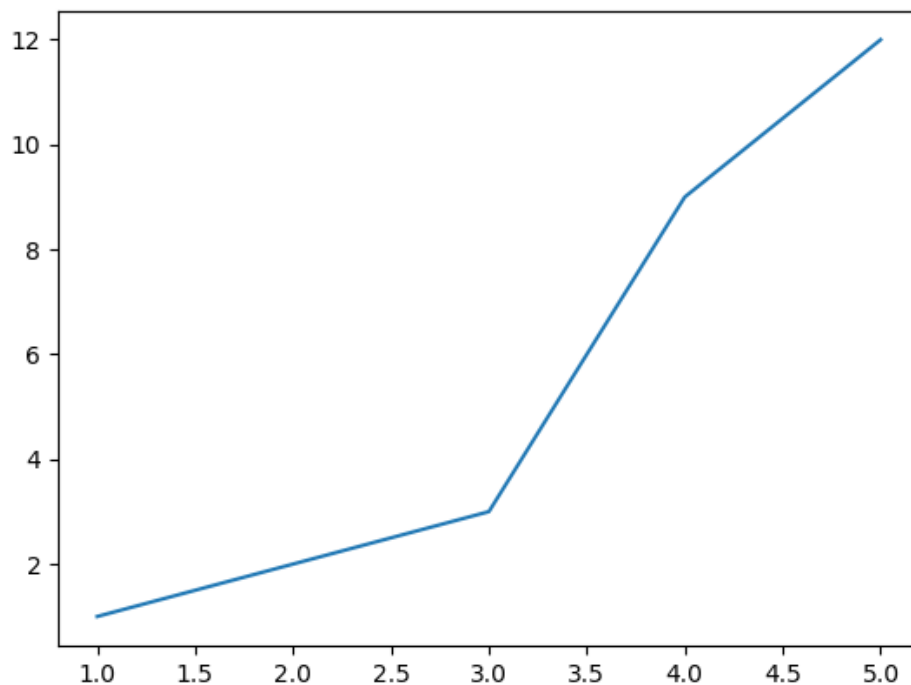
```
>>> Y = [1, 3, 9, 12]
>>> X = range(len(Y))
>>> plt.plot(X, Y)
[<matplotlib.lines.Line2D object at 0x000002042BEB128>]
plt.show()
```

¹To se zgodi v primeru, ko imamo v prvem seznamu številske vrednosti. Če bi imeli v prvem seznamu nize, bi bili ti uporabljeni kot oznake na osi x , pri čemer bi bile lokacije točk na osi x zopet določene kar z indeksi točk na osi y .

Os x bi lahko tudi spremenili. Poskusimo:

```
>>> X = [1,3,4,5]
>>> Y = [1, 3, 9, 12]
>>> plt.plot(X,Y)
[<matplotlib.lines.Line2D object at 0x000002042A9D0DA0>]
>>> plt.show()
```

Graf, ki smo ga narisali tako, prikazuje slika 14.2. Na sliki je prikazan samo zadnji



Slika 14.2 Črtni graf, pri čemer smo podali koordinate točk na obeh oseh.

graf. Kam je izginil prejšnji? Kot smo že omenili knjižnica Matplotlib grafe riše na risalni površini v ozadju. Te prikaže, ko pokličemo funkcijo `show`, hkrati pa takrat risalno površino tudi počisti. Če hočemo na isti sliki prikazati več grafov, bomo pred klicanjem funkcije `show` pač narisali več grafov. Poskusimo to kar na zgledu s plačami, in sicer bi radi narisali podatke o povprečnih bruto plačah. Predpostavljali bomo, da smo funkcijo `uvozi_place` shranili v program `place_beri.py` in da se ta program nahaja v naši trenutni delovni mapi. Najprej bomo uvozili funkcijo za uvoz podatkov o plačah zraven pa še knjižnico Matplotlib:

```
>>> from place_beri import uvozi_place
>>> import matplotlib.pyplot as plt
```

Potem preberimo podatke o plačah:

```
>>> place = uvozi_place('place.csv')
```

in potegnemo sezname is slovarjev:

```
>>> MJ = place['Javni sektor']['mesec']
>>> ZJ = place['Javni sektor']['bruto']
>>> MZ = place['Zasebni sektor']['mesec']
>>> ZZ = place['Zasebni sektor']['bruto']
```

Zdaj bomo kot koordinate na osi x podali podatke o mesecih. Tako se nam bodo na osi x izpisali kar podatki o mesecih. Kot koordinate na osi y podamo podatke o zneskih. Potem bomo poklicali še funkcijo za prikaz grafa.

```
>>> plt.plot(MJ, ZJ)
>>> plt.plot(MZ, ZZ)
>>> plt.show()
```

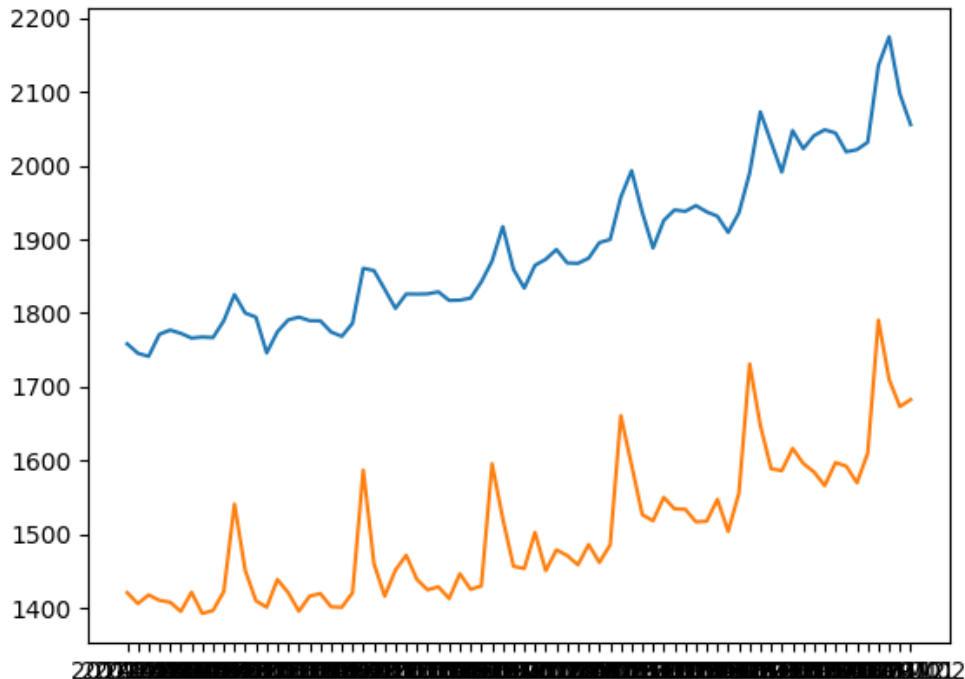
Več grafov lahko na isto sliko narišemo tudi tako, da naštejemo pare seznamov kar po vrsti. Takole:

```
>>> plt.plot(MJ, ZJ, MZ, ZZ)
>>> plt.show()
```

Kljub temu, da je rezultat v zgornjih dveh primerih enak, bomo raje uporabljali prvi način. Zapišimo zdaj vse skupaj kot program `risi_place.py`.

```
1 from place_beri import uvozi_place # funkcija za uvoz
2 import matplotlib.pyplot as plt
3
4 # uvozi podatke v slovar
5 place = uvozi_place('place.csv')
6 # pridobi sezname iz slovarja
7 MJ = place['Javni sektor']['mesec']
8 ZJ = place['Javni sektor']['bruto']
9 MZ = place['Zasebni sektor']['mesec']
10 ZZ = place['Zasebni sektor']['bruto']
11
12 plt.plot(MJ, ZJ) # riši javni sektor
13 plt.plot(MZ, ZZ) # riši zasebni sektor
14 plt.show() # prikaži graf
```

Rezultat izvedbe zgornjega programa prikazuje slika 14.3.



Slika 14.3 Osnovni izris podatkov o plačah.

14.3 Dodajanje oznak

Vsak graf seveda potrebuje oznake. Označiti moramo kaj prikazuje posamezna os, za kar lahko uporabimo funkciji `xlabel` in `ylabel`, ki kot argument prejmeta niz, ki ga želimo prikazati. V našem primeru bi bilo smiselno napisati takole:

```
plt.xlabel("mesec")
plt.ylabel("znesek [EUR]")
```

Dodamo lahko tudi naslov grafa z uporabo funkcije `title`. Takole:

```
plt.title("Povprečne mesečne plače")
```

Manjka seveda tudi legenda. Kaj prikazuje modra linija in kaj oranžna? Legendo lahko dodamo tako, da oznake dodamo posameznemu grafu kar med izrisom. Funkciji `plot` lahko preko opcijskega argumenta `label` podamo niz, ki predstavlja oznako grafa, ki ga bo izrisala. V našem primeru bi to naredili takole:

```
plt.plot(MJ, ZJ, label='Javni sektor') # riši javni sektor
plt.plot(MZ, ZZ, label='Zasebni sektor') # riši zasebni sektor
```

Če želimo legendo pokazati, moramo poklicati še funkcijo `legend`, ki prikaz legende vklopi:

```
plt.legend()
```

Vsebinsko legende, ki jo želimo izpisati bi lahko podali tudi neposredno funkciji `legend`. Takole:

```
plt.legend(['Javni sektor', 'Zasebni sektor'])
```

Pri tem moramo paziti na to, da oznake v legendi podajamo v enakem vrstnem redu, kot smo izvajali risanje grafov.

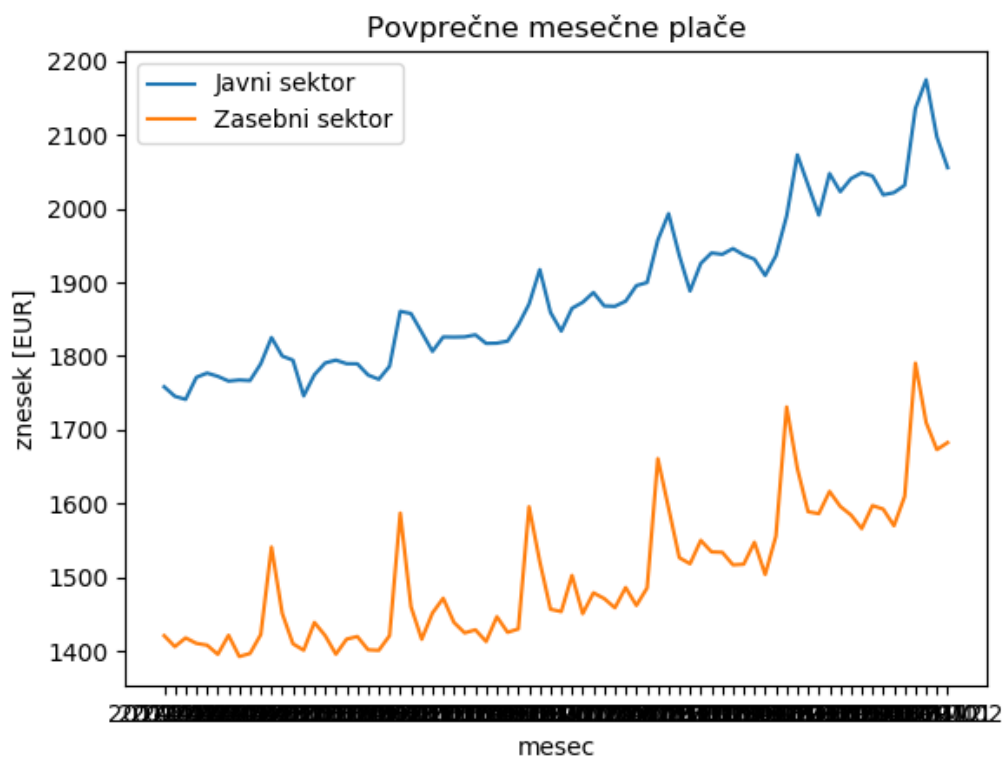
Zapišimo celoten program, ga poženimo in pogledjmo rezultat.

```
1 from place_beri import uvozi_place # funkcija za uvoz
2 import matplotlib.pyplot as plt
3
4 # uvozi podatke v slovar
5 place = uvozi_place('place.csv')
6 # pridobi sezname iz slovarja
7 MJ = place['Javni sektor']['mesec']
8 ZJ = place['Javni sektor']['bruto']
9 MZ = place['Zasebni sektor']['mesec']
10 ZZ = place['Zasebni sektor']['bruto']
11
12 plt.plot(MJ, ZJ) # riši javni sektor
13 plt.plot(MZ, ZZ) # riši zasebni sektor
14
15 # dodaj oznake
16 plt.xlabel("mesec")
17 plt.ylabel("znesek [EUR]")
18 plt.title("Povprečne mesečne plače")
19 plt.legend(['Javni sektor', 'Zasebni sektor'])
20
21 plt.show() # prikaži graf
```

Rezultat izvedbe zgornjega programa prikazuje slika 14.4.

14.4 Še malo prilagajanja oznak

Kar nam še vedno ni všeč na sliki 14.4, je neberljiv izpis na osi x . Če graf približamo, vidimo, da je na oseh izpisan podatek o mesecu. Mogoče bi bilo bolje, če bi ta podatek izpisali samo vsak januar, poleg tega pa bi bilo potem smiselno izpisati samo informacijo o letu (brez meseca). Poskusimo odrezati rezino po mesecih od začetka do konca, pri čemer za korak nastavimo vrednost 12.



Slika 14.4 Izris podatkov o plačah z dodanimi oznakami.

```
>>> MJ[::12]
['2014M01', '2015M01', '2016M01', '2017M01', '2018M01',
 '2019M01', '2020M01']
```

Pripravimo si seznam *oznake*, ki bo vseboval samo podatke o letih. Vzeli bomo vsak 12-ti podatek iz obstoječega seznama mesecev (izhajamo lahko bodisi iz seznama MJ ali MZ), pri čemer bomo upoštevali samo prve štiri znake (podatek o letu).

```
oznake = []
for mj in MJ[::12]: # vzamemo vsako 12-to oznako
    oznake.append(mj[:4]) # vzamemo samo podatek o letu
```

Določiti moramo še lokacije, kjer bomo te oznake prikazali. Trenutno so oznake prikazane na lokacijah, ki se ujemajo z njihovimi indeksi, torej bi lahko lokacije oznak dobili s seznamoma `range(len(MJ))` ter `range(len(MZ))`. Ker bi radi prikazali vsako 12-to oznako, bomo morali torej upoštevati tudi vsako 12-to lokacijo. Takole:

```
# vsaka 12-ta lokacija
```

```
lokacije = range(0, len(MJ), 12)
```

Lokacijo in vsebino oznak lahko zdaj našemu risarju podamo preko funkcije `xticks` (če bi želeli prilagajati oznake na osi *y*, bi uporabili funkcijo `yticks`):

```
plt.xticks(lokacije, oznake)
```

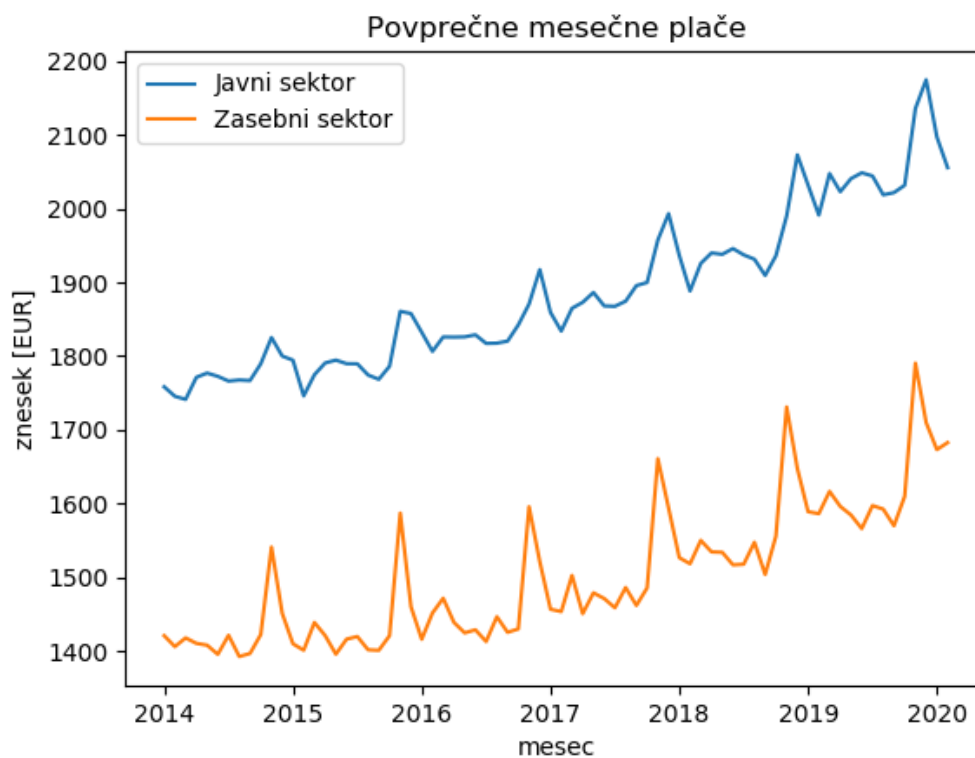
Celoten program je zdaj sledeč:

```
1 from place_beri import uvozi_place # funkcija za uvoz
2 import matplotlib.pyplot as plt
3
4 # uvozi podatke v slovar
5 place = uvozi_place('place.csv')
6 # pridobi sezname iz slovarja
7 MJ = place['Javni sektor']['mesec']
8 ZJ = place['Javni sektor']['bruto']
9 MZ = place['Zasebni sektor']['mesec']
10 ZZ = place['Zasebni sektor']['bruto']
11
12
13 # dodaj oznake
14 plt.xlabel("mesec")
15 plt.ylabel("znesek [EUR]")
16 plt.title("Povprečne mesečne plače")
17 plt.legend(['Javni sektor', 'Zasebni sektor'])
18
19 oznake = []
20 for mj in MJ[::12]: # vzamemo vsako 12-to oznako
21     oznake.append(mj[:4]) # vzamemo samo podatek o letu
22
23 # vsaka 12-ta lokacija
24 lokacije = range(0, len(MJ), 12)
25
26 plt.xticks(lokacije, oznake)
27
28 plt.show() # prikaži graf
```

Rezultat izvedbe programa prikazuje slika 14.5.

Kaj pa če bi imeli v podatkih o plačah še kakšen sektor več? Ker imamo podatke shranjene v dokaj prilagodljivi strukturi (slovarju), bi lahko izris naredili neodvisno od števila sektorjev. Enostavno se sprehodimo čez ključne slovarja in rišemo. Takole:

```
for sektor in place:
    mesec = place[sektor]['mesec']
```



Slika 14.5 Izris podatkov o plačah s prilagojenimi oznakami na osi x .

```
znesek = place[sektor]['bruto']
plt.plot(mesec, znesek, label=sektor)
```

Tokrat smo oznake grafov dodajali že kar med izrisovanjem preko argumenta `label`. Pri risanju oznak na osi x lahko uporabimo kar spremenljivko `mesec`, v kateri so ostali podatki zadnjega sektorja (po sprehodu z zanko `for`). Celotna koda je sledeča:

```
1 from place_beri import uvozi_place # funkcija za uvoz
2 import matplotlib.pyplot as plt
3
4 # uvozi podatke v slovar
5 place = uvozi_place('place.csv')
6 # pridobi sezname iz slovarja
7 MJ = place['Javni sektor']['mesec']
8 ZJ = place['Javni sektor']['bruto']
9 MZ = place['Zasebni sektor']['mesec']
10 ZZ = place['Zasebni sektor']['bruto']
```

```

11
12
13 plt.plot(MJ, ZJ) # riši javni sektor
14 plt.plot(MZ, ZZ) # riši zasebni sektor
15
16 # dodaj oznake
17 plt.xlabel("mesec")
18 plt.ylabel("znesek [EUR]")
19 plt.title("Povprečne mesečne plače")
20 plt.legend(['Javni sektor', 'Zasebni sektor'])
21
22 oznake = []
23 for mj in MJ[::12]: # vzamemo vsako 12-to oznako
24     oznake.append(mj[:4]) # vzamemo samo podatek o letu
25
26 # vsaka 12-ta lokacija
27 lokacije = range(0, len(MJ), 12)
28
29 # prikaži oznake na osi x
30 plt.xticks(lokacije, oznake)
31
32 plt.show() # prikaži graf

```

14.5 Ostale prilagoditve izrisa

Če nam grafi še vedno niso všeč, se lahko igramo naprej. Preko funkcije `plot` lahko nastavljamo barvo izrisa (argument `color`), debelino črte (argument `linewidth`), tip črte (argument `linestyle`) in še marsikaj. Poleg tega lahko določamo razpon osi (funkcija `axis`), rišemo več podgrafov (funkcija `subplot`) in graf shranjujemo v datoteko (funkcija `savefig`). Možnosti je res veliko in jih tukaj ne bomo več naštevali. Primere različnih grafov, ki jih lahko izrišemo z uporabo knjižnice Matplotlib, si lahko bralec pogleda (in prosto dostopno kodo prilagodi za risanje svojih grafov) na povezavi <https://matplotlib.org/gallery>.

14.6 Ostali tipi grafov

Matplotlib poleg črtnega diagrama (funkcija `plot`) omogoča risanje tudi ostalih tipov grafov, npr. stolpčnega diagrama (angl. *bar plot*) s funkcijo `bar`, histograma s funkcijo `hist`, kvartilnega diagrama s funkcijo `box` itd. Poglejmo si še primer izrisa stolpčnega diagrama, pri čemer bomo prikazali podatke o povprečnih plačah za leto 2018. Najprej iz podatkov izluščimo zgolj podatke za leto 2018. Hkrati se

bomo morali sprehajati čez mesece in zneske. Ker sta seznama poravnana (isti indeks se nanaša na isti mesec), lahko naredimo sprehod s pomočjo funkcije `zip`. Znotraj sprehoda pogledamo, če se mesec nanaša na leto 2018 in v tem primeru mesec in znesek dodamo v nova seznama, ki se nanašata na leto 2018. To naredimo za javni in zasebni sektor posebej:

```
# uvozi podatke v slovar
place = uvozi_place('place.csv')
# pridobi sezname iz slovarja
MJ = place['Javni sektor']['mesec']
ZJ = place['Javni sektor']['bruto']
MZ = place['Zasebni sektor']['mesec']
ZZ = place['Zasebni sektor']['bruto']

MJ_2018 = []
ZJ_2018 = []
MZ_2018 = []
ZZ_2018 = []

for mj, zj in zip(MJ, ZJ):
    if "2018" in mj:
        MJ_2018.append(mj)
        ZJ_2018.append(zj)

for mz, zz in zip(MZ, ZZ):
    if "2018" in mz:
        MZ_2018.append(mz)
        ZZ_2018.append(zz)
```

Zdaj lahko podatke narišemo, pri čemer bomo za izris uporabili funkcijo `bar`. Ena izmed razlik med funkcijo `plot` in `bar` je, da moramo pri slednji lokacije stolpcev na osi x vedno podati. Uporabimo lahko kar funkcijo `range`:

```
plt.bar(range(len(ZJ_2018)), ZJ_2018)
plt.bar(range(len(ZZ_2018)), ZZ_2018)
```

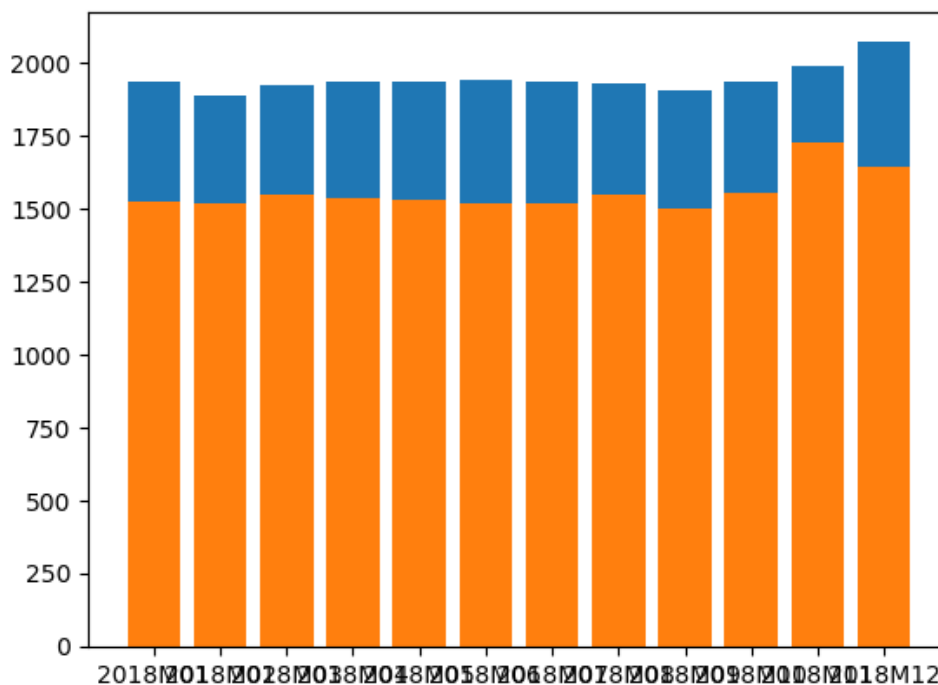
Z uporabo funkcije `xticks` lahko določimo še oznake na osi x :

```
plt.xticks(range(len(MJ)), MJ)
```

Poglejmo si rezultat

```
plt.show()
```

Prikazuje ga slika 14.6. Oznake na osi x so zopet moteče. Grafu bi lahko dodali naslov, da gre za leto 2018, na osi x pa prikazali samo informacijo o mesecu. Tako bo postal graf bolj pregleden. Najprej iz oznak odstranimo podatek



Slika 14.6 Izris podatkov o plačah za leto 2018 s stolpčnim diagramom.

o letu. To lahko naredimo že med filtriranjem podatkov za leto 2018, kjer namesto stavkov `MJ_2018.append(mj)` in `MZ_2018.append(mz)` uporabimo stavka `MJ_2018.append(mj[-3:])` in `MZ_2018.append(mz[-3:])`. Naslov grafa dodamo s sledečo vrstico:

```
plt.title('Podatki o plačah za leto 2018')
```

Dodajmo še legendo:

```
plt.legend(['Javni sektor', 'Zasebni sektor'])
```

Kar nas še moti je to, da sta grafa *naložena* drug na drugega. Če bi npr. najprej izrisali zasebni sektor, bi ga javni sektor prekril. Bolj pravilno bi bilo, če bi grafa narisala drug ob drugemu. Kako lahko to naredimo? Tako, da en graf zamaknemo v levo, drugega pa v desno. Popravili bomo tudi širino stolpcev, da bodo malo ožji. Najprej določimo lokacije stolpcev:

```
loc1 = list(range(len(ZJ_2018)))
loc2 = list(range(len(ZZ_2018)))
for i in range(len(loc1)):
```

```
loc1[i] -= 0.2
loc2[i] += 0.2
```

Zdaj te lokacije upoštevajmo pri izrisu, poleg tega pa preko izbirnega argumenta `width` nastavimo še širino stolpcev:

```
plt.bar(loc1, ZJ_2018, width = 0.4)
plt.bar(loc2, ZZ_2018, width = 0.4)
```

Celoten program je zdaj sledeč:

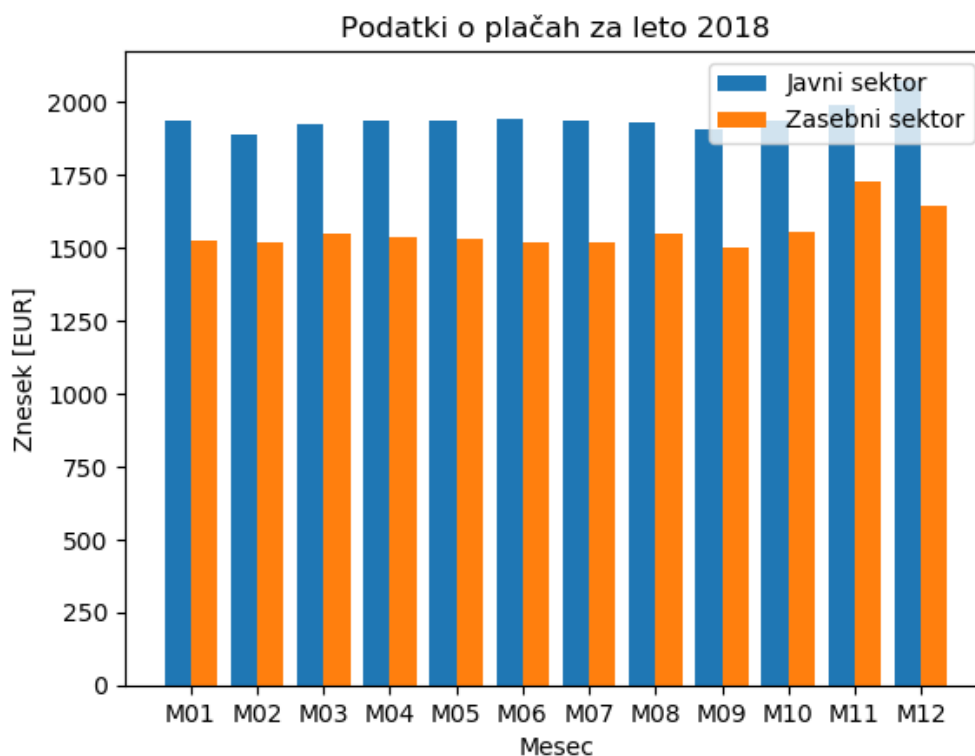
```
1 from place_beri import uvozi_place # funkcija za uvoz
2 import matplotlib.pyplot as plt
3
4 # uvozi podatke v slovar
5 place = uvozi_place('place.csv')
6 # pridobi sezname iz slovarja
7 MJ = place['Javni sektor']['mesec']
8 ZJ = place['Javni sektor']['bruto']
9 MZ = place['Zasebni sektor']['mesec']
10 ZZ = place['Zasebni sektor']['bruto']
11
12 # izluščimo podatke za leto 2018
13 MJ_2018 = []
14 ZJ_2018 = []
15 MZ_2018 = []
16 ZZ_2018 = []
17
18 for mj, zj in zip(MJ, ZJ):
19     if "2018" in mj:
20         MJ_2018.append(mj[-3:]) # brez leta
21         ZJ_2018.append(zj)
22
23 for mz, zz in zip(MZ, ZZ):
24     if "2018" in mz:
25         MZ_2018.append(mz[-3:]) # brez leta
26         ZZ_2018.append(zz)
27
28 loc1 = list(range(len(ZJ_2018)))
29 loc2 = list(range(len(ZZ_2018)))
30 for i in range(len(loc1)):
31     loc1[i] -= 0.2
32     loc2[i] += 0.2
33
```

```

34 plt.bar(loc1, ZJ_2018, width = 0.4)
35 plt.bar(loc2, ZZ_2018, width = 0.4)
36 plt.xticks(range(len(MJ_2018)), MJ_2018)
37
38 plt.title('Podatki o plačah za leto 2018')
39 plt.legend(['Javni sektor', 'Zasebni sektor'])
40 plt.xlabel("Mesec")
41 plt.ylabel("Znesek [EUR]")
42 plt.show()

```

Rezultat izvedbe programa prikazuje slika 14.7.



Slika 14.7 Dopolnjen in popravljen izris podatkov o plačah za leto 2018 s stolpčnim diagramom.

14.7 Risanje matematičnih funkcij

Kaj pa če bi želeli narisati graf matematične funkcije? Podobno kot v zgornjih primerih, bomo funkciji za risanje podali seznam koordinat na osi x in seznam

koordinat na osi y . Koordinate na osi x lahko določimo glede na željen razpon vrednosti – tega določimo sami. Za določitev vrednosti na osi y pa moramo funkcijo tabelirati pri podanih vrednostih osi x . Poskusimo na zgledu:

Zgled 64 *Nariši grafe logaritemskih funkcij z osnovo 2, e in 10 na intervalu od 0 do 5. Sliko opremi z legendami, vsak graf pa naj bo narisani s svojo barvo.*

Rešitev 64 *Najprej bomo določili koordinate osi x . Lahko bi jih našteali, lahko pa uporabimo kakšno funkcijo za generiranje števil v podanem intervalu – npr. funkcijo `range`.*

```
>>> X = range(6)
```

Če želimo v razpon vključiti tudi število 5, moramo kot argument `stop` podati število 6. Kako bomo prišli do funkcij za izračun logaritmov? Spomnimo se na modul `math`. Če pogledamo njegovo dokumentacijo, vidimo, da vsebuje funkcije `log2`, `log` in `log10`. Uvozimo jih.

```
>>> from math import log2, log, log10
```

Poskusimo zdaj izračunati koordinate točk na osi y :

```
>>> Y_2 = log2(X)
```

```
TypeError: must be real number, not range
```

Tole ne bo šlo. funkcije za izračun logaritmov znajo delati samo s številskimi argumenti, mi pa smo podali seznam. Logaritem bomo morali torej izračunati za vsako točko posebej. Poskusimo najprej z dvojiškim logaritmom. Najprej bomo naredili prazen seznam. Potem bomo naredili sprehod čez vse točke na osi x in za vsako posebej izračunali logaritem ter ga dodali v seznam.

```
>>> Y_2 = []
```

```
>>> for x in X:
```

```
    Y_2.append(log2(x))
```

```
ValueError: math domain error
```

Kaj je narobe tokrat? Tokrat problem ni programerski, ampak matematični, saj logaritem števila 0 ni definiran. Razpon točk na osi x bo torej potrebno zmanjšati na interval od vrednosti 1 do 5. Takole:

```
>>> X = range(1,6)
```

```
>>> Y_2 = []
```

```
>>> for x in X:
```

```
    Y_2.append(log2(x))
```

Podobno naredimo še za naravni in desetiški logaritem in stvar narišemo. Celotna rešitev bo sledeča:

```

1  from math import log2, log, log10
2  import matplotlib.pyplot as plt
3
4  X = range(1,6) # koordinate točk na x osi
5
6  Y_2 = []
7  Y_e = []
8  Y_10 = []
9
10 for x in X:
11     Y_2.append(log2(x)) # dvojšiki
12     Y_e.append(log(x)) # naravni
13     Y_10.append(log10(x)) # desetiški
14
15 plt.plot(X, Y_2, label="$log_2(x)$")
16 plt.plot(X, Y_e, label="$log_e(x)$")
17 plt.plot(X, Y_10, label="$log_{10}(x)$")
18 plt.xlabel('x')
19 plt.ylabel('y')
20 plt.legend()
21
22 plt.show()

```

Graf, ki ga na ta način dobimo, prikazuje slika 14.8. Problem je, da ta graf zgolj približno spominja na izris logaritemskih funkcij. Izrisali smo namreč samo 5 točk in te med seboj povezali. Prav tako nam manjka izris točk, ki imajo koordinato x na intervalu $(0, 1)$. Kako bi lahko resolucijo izrisovanja povečali? Tako, da bi zmanjšali korak, s katerim generiramo koordinate na osi x . Idealno bi bilo, če bi izrisovanje začeli pri neki zelo majhni vrednosti x -a (npr. 0.001), potem pa to vrednost povečevali do števila 5, z nekim zelo majhnim korakom (npr. 0.001). Funkciji `range` bi to podali na sledeč način:

```

>>> X = range(0.001, 5.001, 0.001)
TypeError: 'float' object cannot be interpreted as
an integer

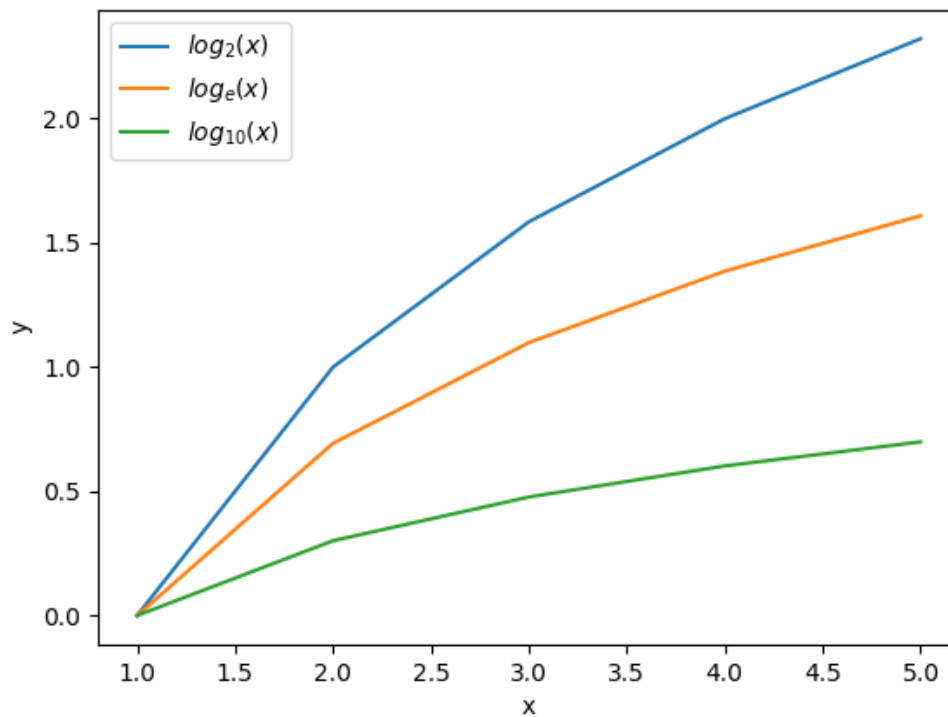
```

Problem tega pristopa je ta, da lahko funkciji `range` podamo samo celoštevilске argumente. Znajti se bomo morali torej drugače. Alternativen pristop bi bil, da s funkcijo `range` zgeneriramo točke od 1 do 5000, in potem vsako izmed njih delimo s 1000. Tako bomo dobili definicijsko območje z željeno resolucijo.

```

>>> X_cel = range(1,5001,1)
>>> X = []
>>> for x in X_cel:

```



Slika 14.8 Osnovni izris logaritemskih funkcij.

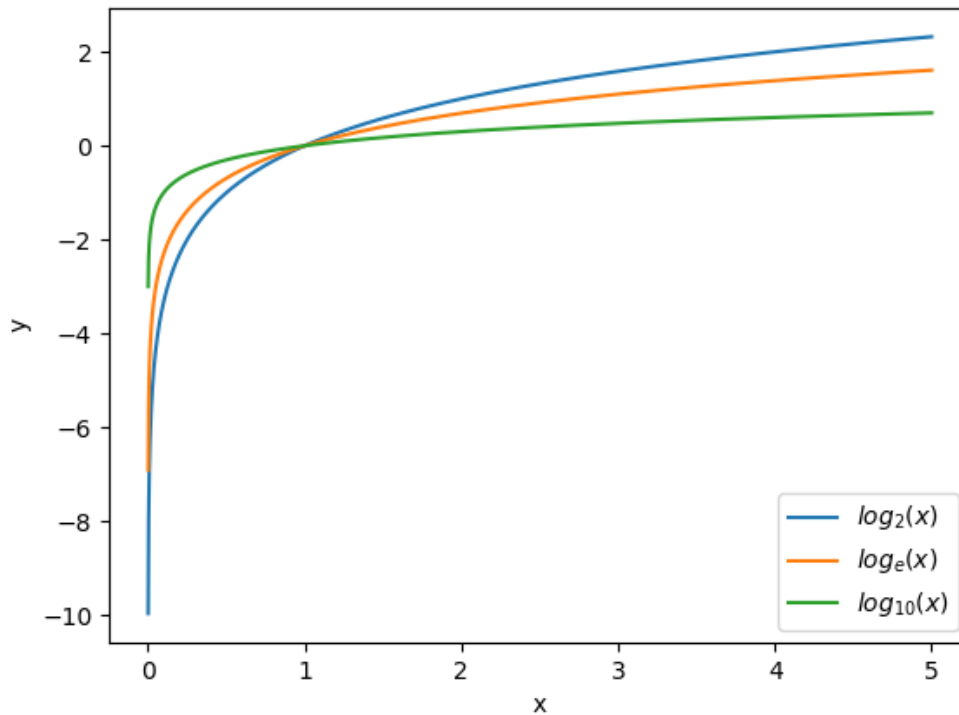
```
x /= 1000
X.append(x)
```

Zdaj lahko pri računanju upoštevamo te vrednosti. Dopolnimo zgornjo kodo in pogledimo rezultat, ki ga prikazuje slika 14.9. Celotno rešitev prikazuje spodnja koda:

```
>>> X = range(1,6)
>>> Y_2 = []
>>> for x in X:
>>>     Y_2.append(log2(x))
```

Podobno naredimo še za naravni in desetiški logaritem in stvar narišemo. Celotna rešitev bo sledeča:

```
1 from math import log2, log, log10
2 import matplotlib.pyplot as plt
3
4 X_cel = range(1,5001,1) # predpriprava za koordinate x
5
```



Slika 14.9 Izris logaritmskih funkcij z ustrežnejšo resolucijo.

```

6  X = []
7  Y_2 = []
8  Y_e = []
9  Y_10 = []
10
11 for x in X_cel:
12     x /= 1000 # koordinata x
13     X.append(x)
14
15     Y_2.append(log2(x)) # dvojšiki
16     Y_e.append(log(x)) # naravni
17     Y_10.append(log10(x)) # desetiški
18
19 plt.plot(X, Y_2, label="$log_2(x)$")
20 plt.plot(X, Y_e, label="$log_e(x)$")
21 plt.plot(X, Y_10, label="$log_{10}(x)$")
22 plt.xlabel('x')

```

```
23 plt.ylabel('y')
24 plt.legend()
25
26 plt.show()
```

Tole je bilo precej nerodno predvsem iz dveh razlogov. Prvič, za izračun logaritmov vseh vrednosti v seznamu **X** smo se morali sprehoditi čez cel seznam in vsako vrednost izračunati posebej. Veliko lažje bi bilo, če bi lahko funkcijo poklicali kar nad celotnim seznamom (brez uporabe zanke `for`). Drugič, funkciji **range** lahko podamo samo celoštevilске argumente. Ker smo želeli biti pri izrisovanju logaritmov nekoliko bolj natančni, smo morali generiranje definicijskega območja nekoliko zakomplicirati. Idealno pa bi bilo, če bi lahko funkciji **range** oziroma njej podobni funkciji podali decimalne korake. Izkaže se, da lahko zgornja problema rešimo z uporabo knjižnice **NumPy**, ki pa nam poleg tega olajša še marsikaj drugega.

Osnutek

15 Knjižnica NumPy in hitro računanje

15.1 NumPy in ndarray

Knjižnica NumPy nadgrajuje Pythonove sezname, poleg tega pa nudi številne funkcije, ki nam olajšajo predvsem matematične operacije nad t.i. nadgrajenimi seznamami oziroma strukturo `ndarray` (angl. *N-dimensional array*). Vse se bo torej vrtelo okrog strukture `ndarray`. Kljub temu, da osnovna namestitev okolja Python knjižnice NumPy ne vsebuje, se je knjižnica namestila skupaj z namestitvijo knjižnice Matplotlib, saj jo slednji potrebuje za svoje delovanje. Uvozimo jo v svoje delovno okolje.

```
>>> import numpy as np
```

Zdaj lahko poljuben seznam (ali seznamu podobno strukturo) pretvorimo v `ndarray` z uporabo funkcije `array`:

```
>>> X = np.array([1,2,3])
>>> Y = np.array([4,5,6])
```

Kaj lahko z *novimi* seznamami počnemo? Podobno kot običajne sezname, lahko tudi te indeksiramo, nad njimi uporabljamo vgrajene funkcije, preverjamo, če vsebujejo določen element in se čez njih sprehajamo z zanko `for`:

```
>>> X[0]
1
>>> len(X)
3
>>> 1 in X
True
>>> for x in X:
        print(x)
1
2
3
```

15.2 Aritmetični operatorji in struktura ndarray

Seznamne smo lahko med seboj tudi seštevali. Tudi ndarray-e lahko, vendar je rezultat nekoliko drugačen:

```
>>> X+Y
array([5, 7, 9])
```

Dobili smo torej nov ndarray, ki je sestavljen iz vsote istoležnih elementov izhodiščnih ndarray-ev. Na tak ndarray lahko torej gledamo kot na vektor, na vsoto dveh ndarray-ev pa kot na vsoto dveh vektorjev. Zdaj je pogoj za seštevanje to, da sta oba vektorja oziroma ndarray-a enako dolga. Tole, npr. ne bo šlo:

```
>>> Z = np.array([7,8])
>>> X+Z
ValueError: operands could not be broadcast together
with shapes (3,) (2,)
```

Lahko pa vektorju prištejemo skalar:

```
>>> X+10
array([11, 12, 13])
```

Nad strukturami tipa ndarray lahko za razliko od seznamov uporabimo tudi druge aritmetične operatorje. Poskusimo:

```
>>> X*Y
array([ 4, 10, 18])
>>> X/Y
array([0.25, 0.4 , 0.5 ])
>>> X-Y
array([-3, -3, -3])
>>> X**5
array([ 1, 32, 243], dtype=int32)
>>> X/10
array([0.1, 0.2, 0.3])
```

Aritmetične operatorje lahko torej na tak način izvedemo nad vsakim elementom vektorja posebej. Poleg tega, da za to ne potrebujemo zank, so te operacije računsko zelo učinkovite in delujejo hitro tudi v primeru, ko delamo z velikimi količinami podatkov.

15.3 Primerjalni operatorji, indeksiranje s seznamami in filtriranje

Podobno kot aritmetični operatorji delujejo nad strukturo ndarray tudi primerjalni operatorji, in sicer tako, da primerjanje izvedejo nad vsakim elementom strukture

posebej. Primerjamo lahko npr. dva vektorja:

```
>>> X = np.array([1,2,3])
>>> Z = np.array([0.1, 10, -2])
>>> X < Z
array([False,  True,  False])
```

Dobili smo torej vektor, ki vsebuje rezultate preverjanja na posameznih mestih. Primerjanje bi lahko izvajali tudi s skalarjem

```
>>> X >= 2
array([False,  True,  True])
```

Dodatna prednost uporabe strukture `ndarray` je, da lahko to indeksiramo z vektorjem vrednosti `True` in `False`, pri čemer bomo kot rezultat dobili `ndarray` elementov, ki smo jih naslovili z vrednostjo `True`. Rezultat primerjanja strukture `ndarray` lahko torej uporabimo za indeksiranje. Če bi npr. želeli dobiti tiste elemente vektorja `X`, ki so večji ali enaki 2, bi lahko to naredili takole:

```
>>> sito = X >= 2
>>> X[sito]
array([2, 3])
```

oziroma krajše:

```
>>> X[X >= 2]
array([2, 3])
```

Tej operaciji lahko rečemo tudi *filtriranje* vrednosti.

15.4 Generiranje strukture `ndarray`

Kot smo videli zgoraj lahko strukturo `ndarray` dobimo tako, da s funkcijo `array` vanjo pretvorimo običajen seznam. Modul NumPy pa ponuja tudi številne funkcije, ki jih lahko uporabimo pri generiranju struktur `ndarray` z vnaprej poznanimi lastnostmi. Strukturo `ndarray`, ki npr. vsebuje same ničle, lahko zgeneriramo s funkcijo `zeros`, same enice s funkcijo `ones`, matriko naključnih števil pa preko modula `numpy.random`. Podrobneje si bomo pogledali funkcijo `arange` in `linspace`, ki sta namenjeni generiranju struktur `ndarray` na vnaprej določenem intervalu. Funkcija `arange` deluje zelo podobno kot funkcija `range`, le da vrača strukturo tipa `ndarray` poleg tega pa za razliko od funkcije `range` ni omejena na celoštevilске argumente. Če bi npr. želeli zgenerirati vektor vrednosti od 0.001 do 5 s korakom 0.001, bi to naredili s sledečim klicem:

```
>>> X = np.arange(0.001, 5.001, 0.001)
```

Funkcija `linspace` deluje podobno, le da tej kot argumente podamo začetno in končno točko intervala (tokrat bo slednja v intervalu vsebovana) in število točk,

ki jih želimo v intervalu imeti. Enak rezultat, kot ga dobimo z gornjim klicem funkcije `arange`, lahko dobimo tudi s funkcijo `linspace` takole:

```
>>> X = np.linspace(0.001, 5, 5000)
```

Rekli smo torej, da želimo imeti 5000 točk na intervalu, ki se začne s točko 0.001 in konča s točko 5 (ta točka je v intervalu še vsebovana). Pri tem je uporabljena linearna interpolacija med točkami, kar z drugimi besedami pomeni, da je med sosednjima točkama v vektorju razlika vselej enaka.

15.5 Funkcije nad strukturo `ndarray`

Če želimo narisati graf logaritemske funkcije, lahko torej koordinate na osi x enostavneje zgeneriramo s funkcijo `arange`. Zdaj moramo le še izračunati njihove logaritme. Poskusimo:

```
>>> from math import log2, log, log10
>>> log2(X)
TypeError: only size-1 arrays can be converted to
Python scalars
```

Funkcije modula `math` torej nad vektorji ne moremo vektorsko izvajati, ampak se moramo spet zateči k zanki `for`. Izkaže pa se, da knjižnica NumPy vsebuje tudi matematične funkcije, ki zamenjujejo tiste iz modula `math`, poleg tega pa podpirajo vektorski način izvajanja. Nad celotnim vektorjem jih torej lahko izvedemo z enim samim klicem. Poskusimo:

```
>>> np.log2(X)
array([-9.96578428, -8.96578428, -8.38082178, ...,
        2.3213509 , 2.32163953, 2.32192809])
```

Zdaj lahko dokončamo zgled z risanjem logaritemskih funkcij.

Zgled 65 *Nariši grafe logaritemskih funkcij z osnovo 2, e in 10 na intervalu od 0 do 5. Sliko opremi z legendami, vsak graf pa naj bo narisani s svojo barvo.*

Rešitev 65 *Rešitev bo zdaj bistveno krajša, saj bomo vrednosti na osi x generirali s funkcijo `np.arange` (ena vrstica), logaritme pa računali s funkcijami `np.log2`, `np.log` in `np.log10` (tri vrstice). Ostala koda bo ostala enaka, prav tako pa bo enak rezultat, ki je prikazan na sliki 14.9.*

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 # hitro generiranje vrednosti na osi x
5 X = np.arange(0.001, 5.001, 0.001)
```

```

6
7 # računanje logaritmov brez zanke for
8 Y_2 = np.log2(X)
9 Y_e = np.log(X)
10 Y_10 = np.log10(X)
11
12 # nespremenjena koda od prej
13 plt.plot(X, Y_2, label="$log_2(x)$")
14 plt.plot(X, Y_e, label="$log_e(x)$")
15 plt.plot(X, Y_10, label="$log_{10}(x)$")
16 plt.xlabel('x')
17 plt.ylabel('y')
18 plt.legend()
19
20 plt.show()

```

15.6 Več dimenzij

Tekom spoznavanja seznamov smo se srečali tudi s seznamami seznamov oziroma z ugnезdenimi seznamami. Primer takega seznama je sledeč:

```
>>> sez = [[1, 2, 3], [4, 5, 6], [7, 8, 9]]
```

Kaj se zgodi, če tak seznam pretvorimo v strukturo `ndarray`:

```

>>> A = np.array(sez)
>>> A
array([[1, 2, 3],
       [4, 5, 6],
       [7, 8, 9]])

```

Na prvi pogled nič kaj presenetljivega, ampak če smo si lahko strukturo `ndarray`, ki smo jo dobili iz navadnega seznama, interpretirali kot vektor, lahko na strukturo `ndarray`, ki jo dobimo iz ugnезdenih seznamov, gledamo kot na matriko oziroma na dvodimenzionalno strukturo, ki ima podatke zapisane v vrsticah (dimenzija 0) in stolpcih (dimenzija 1). Indeksiramo jo lahko tako kot običajne ugnезdene seznane. Do ugnезdenega podseznama na npr. indeksu 1, lahko pridemo takole:

```

>>> A[1]
array([4, 5, 6])

```

V matrični interpretaciji to predstavlja vrstico 1. Do elementa na indeksu 2 vrstice 1 lahko naprej pridemo takole:

```
>>> A[1][2]
```

```
6
```

Delovanje je bilo do tukaj zelo podobno kot pri običajnih ugnezdenih seznamih, le interpretacija je malo drugačna. Enako bi lahko naredili tudi pri običajnem seznamu

```
>>> sez[1][2]
6
```

Pri delu s strukturo `ndarray` lahko do elementov matrike pridemo tudi tako, da indeksa vrstice in stolpca podamo skupaj in tako indeksiranje izvedemo le enkrat. To pri običajnih seznamih ne gre:

```
>>> A[1,2]
6
>>> sez[1,2]
TypeError: list indices must be integers or slices,
not tuple
```

Pri indeksiranju smo torej najprej podali indeks po ničti dimenziji oziroma indeks vrstice, potem pa indeks po prvi dimenziji oziroma indeks stolpca (nič nas ne omejuje pri tem, da ne bi število dimenzij še povečali – tako bi dobili tenzor). Tako dobljene matrike lahko podobno kot vektorje med seboj tudi npr. seštevamo in jim prištevamo skalarje:

```
>>> B = np.array([[10, 11, 12],
                  [-1, -2, -3],
                  [0.1, 0.2, 0.3]])
>>> A+B
array([[11. , 13. , 15. ],
       [ 3. ,  3. ,  3. ],
       [ 7.1,  8.2,  9.3]])
```

Nad njimi lahko delamo tudi rezine, in sicer preko vsake dimenzije posebej. Če bi želeli npr. iz matrike `A` dobiti vrstice od 0 do 2, in stolpce od 1 do konca, bi to napisali takole:

```
>>> A[:2,1:]
array([[2, 3],
       [5, 6]])
```

Lahko bi dobili tudi specifično vrstico. Vrstico 2, bi npr. dobili takole:

```
>>> A[2,:]
array([7, 8, 9])
```

Z zgornjo vrstico smo povedali, da želimo ničto dimenzijo fiksirati na indeks 2 (vrstica številka 2), dimenzijo 1 pa želimo imeti v celoti (vsi stolpci). To bi lahko sicer napisali tudi takole

```
>>> A[2]
array([7, 8, 9])
```

in bi seveda delovalo tudi nad navadnimi seznamami. Kako bi lahko prišli do posameznega stolpca? Podobno kot prej, le da zdaj fiksiramo indeks stolpca in se sprehajamo čez vse vrstice. Do stolpca številka 2 bi torej prišli takole:

```
>>> A[:,2]
array([3, 6, 9])
```

Tega nad običajnimi seznamami ne moremo (tako enostavno) narediti.

15.7 Ostale uporabne funkcije

Knjižnica NumPy ponuja še vrsto drugih uporabnih funkcij, ki pa jih boste spoznali, če boste knjižnico bolj intenzivno uporabljali. Nekaj jih bomo vseeno omenili.

Za izračun vsote elementov, določitev minimalnega in maksimalnega elementa lahko uporabimo vgrajene funkcije `min`, `max` in `sum`. Te funkcije pa odpovejo, ko ima naša struktura več dimenzij ali pa vsebuje posebne vrednosti (glej razdelek 15.8). V tem primeru lahko uporabljamo funkcije modula NumPy z enakimi imeni. V osnovi te funkcije delujejo nad celotno strukturo (nad vsemi dimenzijami). Takole:

```
>>> A = np.array([[1,2,3],[4,5,6],[7,8,9]])
>>> np.min(A)
1
>>> np.max(A)
9
>>> np.sum(A)
45
```

Včasih želimo določeno operacijo izvesti le nad npr. vrsticami ali stolpci matrike. V tem primeru lahko pri uporabi zgornjih funkcij podamo še vrednost izbirnega argumenta `axis`. Če ta argument nastavimo na vrednost 0, se bomo tekom operacije znebili dimenzije 0 (vrstic), kar pomeni, da bomo dobili rezultat izvedbe operacije po stolpcih.

```
>>> A = np.array([[1,2,3],[4,5,6],[7,8,9]])
>>> np.min(A, axis=0)
array([1, 2, 3])
>>> np.max(A, axis=0)
array([7, 8, 9])
>>> np.sum(A, axis=0)
array([12, 15, 18])
```

Če argument `axis` nastavimo na vrednost 1, se bomo s tem znebili dimenzije 1 (stolpcev), kar pomeni, da bomo dobili minimum, maksimum in vsoto vrstic:

```

>>> A = np.array([[1,2,3],[4,5,6],[7,8,9]])
>>> np.min(A, axis=1)
array([1, 4, 7])
>>> np.max(A, axis=1)
array([3, 6, 9])
>>> np.sum(A, axis=1)
array([ 6, 15, 24])

```

15.8 Posebne vrednosti

Knjižnica NumPy omogoča, da kot števila predstavimo tudi posebne vrednosti. Prva taka vrednost je uporabna predvsem v primerih, ko nam določen podatek manjka. Če npr. beležimo višino, telesno maso in starost oseb, se lahko hitro zgodi, da med podatki kakšen manjka (nekdo npr. ne želi povedati koliko je star). Narobe bi bilo, če bi tak podatek postavili na neko privzeto numerično vrednost (npr. 0), saj bi nam to pokvarilo določene statistike, kot je npr. povprečje. Prav tako velikokrat podatka ne moremo kar izpustiti. Če želimo podatke npr. beležiti v matriki tipa `ndarray`, morajo imeti vse vrstice enako število stolpcev. V tem primeru lahko uporabimo posebno vrednost `nan` (angl. *not a number*), ki predstavlja t.i. *placeholder* in drži prazno mesto, hkrati pa ga lahko pri analizi vrednosti obravnavamo drugače kot ostale vrednosti (npr. ga izpustimo). V primeru, da naši podatki vsebujejo vrednost `nan` in želimo to pri analizah ignorirati, lahko namesto funkcij kot so `min`, `max` in `sum` uporabimo funkcije `nanmin`, `nanmax` in `nansum`.

Poleg vrednosti `nan` knjižnica NumPy vsebuje tudi posebno vrednost `inf`, s katero lahko npr. zapišemo rezultat deljenja z nič ali pa logaritem števila 0. Pri tem sicer dobimo opozorilo (ne pa napake):

```

>>> np.log(0)
RuntimeWarning: divide by zero encountered in log
-inf

```

15.9 Uvažanje vrednosti in omejitve strukture `ndarray`

V prejšnjih poglavjih smo že omenili datoteke CSV. Ker se taka oblika zapisovanja pogosto uporablja tudi za numerične podatke, knjižnica NumPy podpira uvoz tovrstnih datotek preko funkcij `genfromtxt` in `loadtxt`. Ker je knjižnica NumPy namenjena delu s števili, obe funkciji vsebino datoteke poskušata vrniti kot `ndarray` števil. Kadar datoteka vsebuje zgolj numerične podatke (števila) funkciji ustvarita enak rezultat. Pri tem (običajno) funkcijama kot argument `encoding` podamo

kodiranje datoteke¹ kot argument `delimiter` pa ločilo, ki je uporabljeno znotraj datoteke.

Naj ima datoteka z imenom `stevila1.csv` sledečo vsebino:

```
1 180.3,87.3
2 161.5,77.3
3 170,56.5
4 156,55.3
```

Preberemo jo lahko takole:

```
>>> A1 = np.genfromtxt("stevila1.csv",
                        encoding="utf8",
                        delimiter=",")
>>> B1 = np.loadtxt("stevila1.csv",
                    encoding="utf8",
                    delimiter=",")
```

Pri tem v obeh primerih dobimo enak rezultat:

```
>>> A1
array([[180.3,  87.3],
       [161.5,  77.3],
       [170. ,  56.5],
       [156. ,  55.3]])
>>> B1
array([[180.3,  87.3],
       [161.5,  77.3],
       [170. ,  56.5],
       [156. ,  55.3]])
```

Do težav lahko pride, kadar datoteka vsebuje tudi podatke, ki niso številskega tipa. Ena izmed glavnih omejitev strukture `ndarray` je namreč ta, da zahteva, da vsi njeni podatki pripadajo enakemu podatkovnemu tipu. Podatkovni tip strukture lahko preverimo preko atributa² `dtype`.

```
1 >>> A1.dtype # do atributov dostopamo brez oklepajev
2 dtype('float64')
```

Vsi podatki v strukturi `A1` so torej decimalna števila³).

¹Če vemo, da datoteka vsebuje samo števila, kodiranja ni potrebno podajati.

²Atribut je spremenljivka, ki pripada določenemu objektu, podobno kot je metoda funkcija, ki pripada določenemu objektu.

³`float64` predstavlja zapis števil v plavajoči vejici s 64 biti oziroma z dvojno natančnostjo. Knjižnica NumPy pri zapisovanju decimalnih števil uporablja večjo natančnost kot vgrajeni podatkovni tip `float`.

Kaj se torej zgodi, če datoteka vsebuje podatke, ki niso števila. Dopolnimo našo datoteko s stolpcem, ki bo vsebovala imena ljudi, ki seveda niso števila. To shranimo v datoteko `stevila2.csv`:

```
1 Janez ,180.3,87.3
2 Andrej ,161.5,77.3
3 Ana ,170,56.5
4 Katja ,156,55.3
```

Funkcija `genfromtxt` bo imena enostavno pretvorila v vrednosti `nan`, saj to lahko zapiše kot število. S tem bo dobljena struktura še vedno lahko vsebovala števila, s katerimi bomo lahko računali. Poskusimo:

```
>>> A2 = np.genfromtxt("stevila2.csv",
                        encoding="utf8",
                        delimiter=",")

>>> A2
array([[ nan, 180.3,  87.3],
       [ nan, 161.5,  77.3],
       [ nan, 170. ,  56.5],
       [ nan, 156. ,  55.3]])
```

S tem smo informacijo o imenih sicer izgubili, smo pa ohranili podatkovni tip strukture, tako da ta še vedno vsebuje števila

```
>>> A2.dtype
dtype('float64')
```

Kaj pa funkcija `loadtxt`? Ta poskuša na vsak način podatke pretvoriti v števila, zato ob odpiranju take datoteke vrne napako:

```
>>> B2 = np.loadtxt("stevila2.csv",
                    encoding="utf8",
                    delimiter=",")

ValueError: could not convert string to float: 'Janez'
```

Lahko jo nekoliko prelisičimo, da ji naročimo, naj datoteko uvozi kot nize. To naredimo preko opsijskega argumenta `dtype`, ki ga nastavimo na vrednost `str` ali pa vrednost `'U'` (angl. *Unicode string*)⁴.

```
>>> B2 = np.loadtxt("stevila2.csv",
                    encoding="utf8",
                    delimiter=",",
                    dtype=str)
array([[ 'Janez', '180.3', '87.3'],
```

⁴Tako kot za decimalna števila knjižnica NumPy tudi za zapisovanje nizov uporablja svoj podatkovni tip `'U'` (angl. *Unicode string*)


```
['Andrej', '161.5', '77.3'],
['Ana', '170', '56.5'],
['Katja', '156', '55.3']], dtype='<U6')
```

Zdaj smo podatke uspešno uvozili, ampak za ceno tega, da z njimi ne moremo več računati. Če bi želeli do numeričnih vrednosti vseeno priti, lahko odvržemo stolpec 0, ostale vrednosti pa pretvorimo v števila. Takole:

```
>>> imena = B2[:,0]
>>> imena
array(['Janez',
      'Andrej',
      'Ana',
      'Katja'],
      dtype='<U6')
>>> vrednosti = B2[:,1:].astype(float)
>>> vrednosti
array([[180.3,  87.3],
      [161.5,  77.3],
      [170. ,  56.5],
      [156. ,  55.3]])
```

Opomba: funkcija `astype` izhodiščne strukture ne spreminja, ampak vrne strukturo predstavljeno s podanim podatkovnim tipom.

Datoteka bi lahko vsebovala tudi imena stolpcev. V našem primeru bi to izgledalo nekako takole (datoteka `stevila3.csv`):

```
1 ime,višina,teža
2 Janez,180.3,87.3
3 Andrej,161.5,77.3
4 Ana,170,56.5
5 Katja,156,55.3
```

V tem primeru lahko funkciji `genfromtxt` naročimo, naj ločeno uvozi imena stolpcev preko izbirnega argumenta `names`, ki ga nastavimo na vrednost `True`:

```
>>> A3 = np.genfromtxt("stevila3.csv",
                        encoding="utf8",
                        delimiter=",",
                        names=True)
```

Zdaj so imena stolpcev zapisana ločeno znotraj atributa `dtype`.

```
>>> A3
array([(nan, 180.3, 87.3),
      (nan, 161.5, 77.3),
      (nan, 170. , 56.5),
```

```

        (nan, 156. , 55.3)],
        dtype=[('ime', '<f8'),
               ('višina', '<f8'),
               ('teža', '<f8')])
>>> A3.dtype.names
('ime', 'višina', 'teža')
```

Do imen stolpcev torej lahko pridemo, je pa malo nerodno. Pri uporabi funkcije `np.loadtxt` bi lahko postopali podobno kot prej in celotno datoteko uvozili kot `ndarray` nizov:

```

>>> B3 = np.loadtxt("stevila3.csv",
                    encoding="utf8",
                    delimiter=",",
                    dtype='U')

>>> B3
array([['ime', 'višina', 'teža'],
       ['Janez', '180.3', '87.3'],
       ['Andrej', '161.5', '77.3'],
       ['Ana', '170', '56.5'],
       ['Katja', '156', '55.3']], dtype='<U6')
```

S tem se seveda spet odpovemo zmožnostim računanja, saj se z nizi kaj veliko pač računati ne da.

15.10 Omejitve knjižnice NumPy

Knjižnica NumPy je torej idealna za delo z numeričnimi podatki. Do problema pa pride, kadar naši podatki poleg numeričnih vrednosti vsebujejo tudi nize, ki jih ne moremo obravnavati kot števila. To so lahko zgolj imena stolpcev ali pa tudi drugi podatki. Pri plačah smo se npr. srečali s podatki o mesecu, v zgornjih primerih pa s podatki o imenu osebe. Ko pridemo do takih podatkov moramo skleniti kompromis. Lahko se jim odpovemo z njihovo pretvorbo v `np.nan`, kar ponavadi ne pride v upošteev. Lahko vse podatke zapišemo v obliki nizov, s čimer se odpovemo zmožnostim računanja. Možna alternativa je tudi, da numerične podatke shranimo v drugo strukturo kot nenumerične. To nam načeloma povzroči dodatno delo, je pa seveda izvedljivo. V vsakem primeru so komplikacije neizogibne. Lahko pa v takem primeru uporabimo knjižnico Pandas, ki je namenjena delu z večjimi količinami podatkov, omogoča učinkovito predstavitev različnih podatkovnih tipov in ne izključuje uporabe knjižnice NumPy pri hitrem računanju, temveč to dopolnjuje.

16 Knjižnica pandas

16.1 Delo s podatki

Podatkovne vede (angl. *data sciences*), ki se ukvarjajo z manipulacijo in analizo podatkov, so v današnjem času postale nepogrešljive na različnih področjih ne samo znanosti, temveč tudi različnih vej gospodarstva (npr. ciljno oglaševanje, organizacija dela, planiranje procesov) in negospodarstva (npr. zdravstvo in medicina). Podatkovne vede namreč na podlagi zajema in analize (večjih količin) podatkov nudijo podporo procesom odločanja, s čimer lahko razpoložljive vire izkoristimo bolj učinkovito. Kdor zna delati s podatki ima danes službo zagotovljeno, še posebej, če zna ta znanja uporabiti na svojem primarnem strokovnem področju, kot je npr. kemija.

Če za bolj resno analizo in manipulacijo podatkov uporabljamo programski jezik Python, bomo slej ko prej pristali na uporabi knjižnice pandas (angl. *Python Data Analysis Library*), ki predstavlja moderno in hitro orodje za delo z večjimi količinami podatkov. Delo s knjižnico pandas je namreč v današnjem času postalo sinonim za delo z večjimi količinami podatkov v jeziku Python. V tem poglavju si bomo pogledali par primerov uporabe knjižnice pandas, ki zgolj nakazujejo prednosti uporabe te knjižnice.

16.2 Knjižnica pandas in dataframe

Za uporabo moramo knjižnico pandas najprej namestiti z orodjem `pip`, kar že znamo. V ukazni vrstici svojega operacijskega sistema napišemo:

```
> pip install pandas
```

Zdaj lahko knjižnico uvozimo v svoj program oziroma v svoje delovno okolje, ponavadi pod psevdonimom `pd`:

```
>>> import pandas as pd
```

Večje količine podatkov bomo ponavadi brali iz datotek zapisanih v obliki CSV. Pandas tako branje omogoča preko funkcije `read_csv`. Poskusimo kar na našem zgledu s plačami.

```
>>> pd.read_csv('place.csv')
UnicodeDecodeError: 'utf-8' codec can't decode byte
0xe8 in position 6: invalid continuation byte
```

Tole ni delovalo, ker moramo podati pravilno dekodiranje. Nastavimo argument `encoding`, poleg tega pa rezultat branja shranimo v spremenljivko:

```
>>> df = pd.read_csv('place.csv', encoding="cp1250")
```

Funkcija `read_csv` vrača prebrane podatke v obliki strukture `dataframe`, ki predstavlja tabelo z zapisanimi podatki. Prvih pet vrstic tabele lahko dobimo z metodo `head()`. Poglejmo kaj smo prebrali.

```
>>> df.head()
      Povprečne mesečne bruto in neto plače pri pravnih
      osebah javnega in zasebnega sektorja, Slovenija,
      mesečno
0  MESEC\t"SEKTOR"\t"Bruto plača Plača za mesec[E...
1           2014M01\t"Javni sektor"\t1758.50\t1151.30
2           2014M01\t"Zasebni sektor"\t1421.34\t932.14
3           2014M02\t"Javni sektor"\t1745.63\t1136.41
4           2014M02\t"Zasebni sektor"\t1406.47\t922.19
```

Tabela še vedno ni *tabela*. Smiselno bi bilo, da prve vrstice z opisom vsebine datoteke, izpustimo (to smo naredili tudi prej), tako da izbirnemu argumentu `skiprows` priredimo vrednost 2. Zakaj bomo tokrat izpustili 2 vrstici, ko smo datoteko brali z metodo `read` pa smo izpustili 3? Struktura `dataframe` bo prvo prebrano vrstico (za izpuščenima dvema vrsticama) uporabila kot glavo tabele. Druga stvar, ki jo moralo našemu bralniku nastaviti je še ločilo oziroma separator, ki je v tem primeru tabulator oziroma znak `'\t'`. Tega nastavimo preko argumenta `sep`. Poskusimo datoteko še enkrat prebrati:

```
>>> df = pd.read_csv('place.csv',
                      encoding="cp1250",
                      skiprows=2,
                      sep='\t')
```

Preverimo, katere stolpce imamo v tabeli

```
>>> df.columns
Index(['MESEC', 'SEKTOR', 'Bruto plača Plača za
mesec[EUR]', 'Neto plača Plača za mesec[EUR]'],
      dtype='object')
```

Na prvi pogled izgleda, da smo tabelo zdaj uspešno uvozili. Stolpce lahko tudi preimenujemo v kaj krajšega:

```
>>> df.columns = ['mesec', 'sektor', 'bruto', 'neto']
```

Zdaj pa izpišimo prvih pet vrstic tabele:

```
>>> df.head()
   mesec  sektor      bruto  neto
0  2014M01  Javni sektor  1758.50  1151.30
1  2014M01  Zasebni sektor  1421.34  932.14
2  2014M02  Javni sektor  1745.63  1136.41
3  2014M02  Zasebni sektor  1406.47  922.19
4  2014M03  Javni sektor  1741.44  1133.47
```

Preverimo lahko tudi strukturo tabele preko atributa `shape`:

```
>>> df.shape
(148, 4)
```

Naša tabela ima torej 148 vrstic in 4 stolpce. Do osnovne statistike lahko pridemo preko metode `describe`

```
>>> df.describe()
      bruto      neto
count  148.000000  148.000000
mean   1694.473919  1099.702230
std     217.336086   133.755268
min    1393.050000   915.740000
25%    1471.885000   963.452500
50%    1743.535000  1136.695000
75%    1865.522500  1203.942500
max    2174.570000  1408.770000
```

Ta nam vrne osnovno statistiko, ampak zgolj za numerične stolpce. Težava je le v tem, da imamo zdaj javni in zasebni sektor združena skupaj. Tudi to bomo rešili v kratkem.

16.3 Indeksiranje tabel

Tabele indeksiramo podobno kot sezname, le da tokrat indeksiranje izvajamo po stolpcih. Do stolpca z oznako `bruto` bi torej lahko prišli takole:

```
>>> df['bruto']
0      1758.50
1      1421.34
2      1745.63
...
146    2055.48
147    1682.86
Name: bruto, Length: 148, dtype: float64
```

Lahko bi dostopali tudi do več stolpcev naenkrat, tako da pri indeksiranju podamo seznam stolpcev:

```
>>> df[['bruto', 'neto']]
      bruto    neto
0    1758.50  1151.30
1    1421.34   932.14
2    1745.63  1136.41
...
146  2055.48  1327.33
147  1682.86  1098.04
[148 rows x 2 columns]
```

Kaj pa, če želimo priti do določenih vrstic? V tem primeru uporabimo metodo `loc`, ki ji znotraj oglatih oklepajev podamo oznako oziroma `index` vrstice. Ko smo prej izpisali tabelo, se je pred stolpcem `mesec` izpisal dodaten stolpec oznak. Ko tabelo preberemo, ima ta stolpcev vrednosti od 0 do števila vrstic – 1, lahko pa `index` nastavimo tudi na kaj drugega. Do 0-te vrstice bi torej prišli takole:

```
>>> df.loc[0]
mesec      2014M01
sektor    Javni sektor
bruto      1758.5
neto       1151.3
Name: 0, dtype: object
```

Metodi `loc` lahko podamo tudi stolpec ali seznam stolpcev:

```
>>> df.loc[0, 'bruto']
1758.5
```

Stolpec oznak vrstic lahko spremenimo tudi na kakšnega izmed obstoječih stolpcev. Naredimo novo tabelo, ki bo imela za `index` kar stolpec `mesec`:

```
>>> df2 = df.set_index('mesec')
>>> df2.head()
      sektor    bruto    neto
mesec
2014M01  Javni sektor  1758.50  1151.30
2014M01  Zasebni sektor  1421.34   932.14
2014M02  Javni sektor  1745.63  1136.41
2014M02  Zasebni sektor  1406.47   922.19
2014M03  Javni sektor  1741.44  1133.47
```

Zdaj bi do vrstico metodi `loc` podali kar `mesec`, ki nas zanima:

```
>>> df2.loc['2018M01']
```

	sektor	bruto	neto
mesec			
2018M01	Javni sektor	1936.41	1246.69
2018M01	Zasebni sektor	1526.94	997.06

Zadnji način indeksiranja, ki je neodvisen od oznak vrstic in stolpcev, uporablja metodo `iloc`. Tej podamo številko vrstice lahko pa tudi stolpca. Takole:

```
>>> df.iloc[1]
mesec          2014M01
sektor      Zasebni sektor
bruto          1421.34
neto           932.14
Name: 1, dtype: object
>>> df.iloc[1,3]
932.14
```

Seveda lahko s to metodo delamo tudi rezine. Če nas zanimajo npr vse vrstice, stolpca številka 2, bi to napisali takole:

```
>>> df.iloc[:,2]
0      1758.50
1      1421.34
2      1745.63
...
146     2055.48
147     1682.86
Name: bruto, Length: 148, dtype: float64
```

16.4 Filtriranje vrednosti

Vrednosti lahko filtriramo na podoben način kot pri uporabi strukture `ndarray`. Lahko bi npr. pogledali samo tiste vrstice, ki pripadajo javnemu sektorju:

```
>>> df['sektor'] == 'Javni sektor'
0      True
1     False
2      True
...
146     True
147     False
Name: sektor, Length: 148, dtype: bool
```

Tako dobljen rezultat primerjanja lahko zdaj uporabimo pri indeksiranju:

```
>>> df[df['sektor'] == 'Javni sektor']
mesec      sektor      bruto      neto
0      2014M01  Javni sektor  1758.50  1151.30
2      2014M02  Javni sektor  1745.63  1136.41
4      2014M03  Javni sektor  1741.44  1133.47
...
144     2020M01  Javni sektor  2096.96  1351.52
146     2020M02  Javni sektor  2055.48  1327.33
[74 rows x 4 columns]
```

Če bi nas zanimala statistika po sektorjih, lahko najprej iz tabele izluščimo posamezen sektor, potem pa nad tem pokličemo metodo `describe`:

```
>>> df[df['sektor'] == 'Javni sektor'].describe()
      bruto      neto
count  74.000000  74.000000
mean   1886.850676 1218.003243
std     106.509594   63.695605
min     1741.440000 1133.470000
25%     1794.557500 1161.612500
50%     1866.195000 1204.535000
75%     1944.307500 1249.582500
max     2174.570000 1408.770000
>>> df[df['sektor'] == 'Zasebni sektor'].describe()
      bruto      neto
count  74.000000  74.000000
mean   1502.097162  981.401216
std     93.494044   59.959989
min     1393.050000  915.740000
25%     1421.515000  931.450000
50%     1471.790000  962.165000
75%     1569.237500 1018.832500
max     1790.520000 1177.510000
```

16.5 Risanje grafov

Tudi risanje grafov postane zelo enostavno. Pokličemo lahko kar metodo `plot`, ki pripada tabeli `dataframe`:

```
>>> df.plot()
<matplotlib.axes._subplots.AxesSubplot object at
0x0000029D25BFF7B8>
```

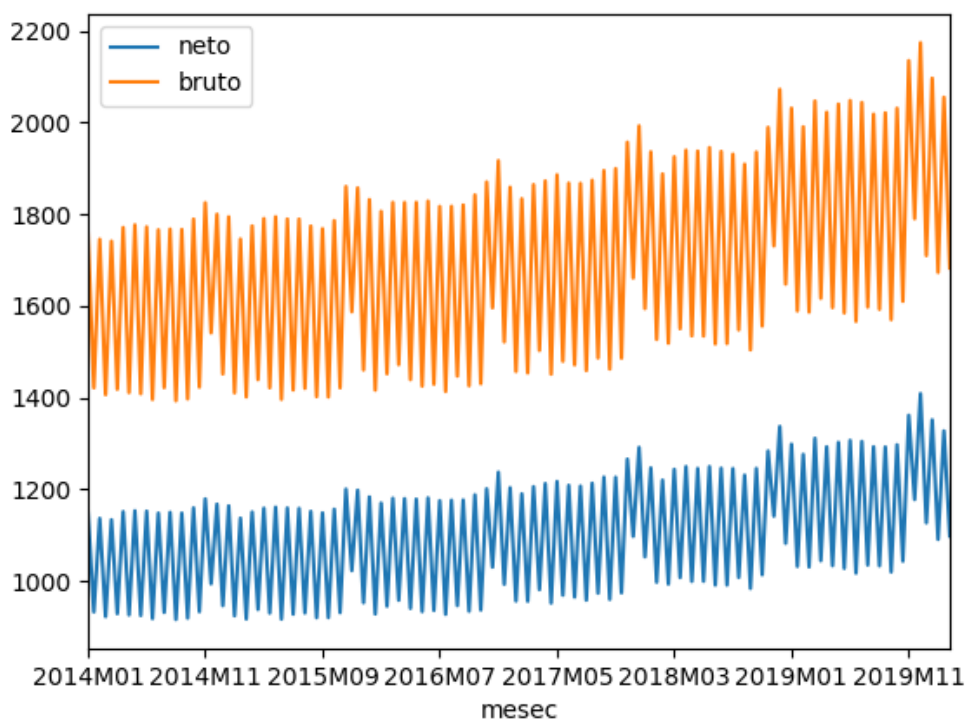

Da lahko tak graf prikažemo, moramo uvoziti še knjižnico Matplotlib in poklicati funkcijo `show`:

```
>>> import matplotlib.pyplot as plt
>>> plt.show()
```

Lahko tudi eksplicitno zahtevamo, kaj naj se prikaže na osi *x*, kaj pa na *y*:

```
df.plot(x='mesec', y=['neto', 'bruto'])
plt.show()
```

Tako dobljen graf prikazuje slika 16.1. Problem dobljenega grafa je, da z isto linijo



Slika 16.1 Črtni graf nad celotno tabelo.

prikazuje javni in zasebni sektor skupaj. Sektorja moramo torej ločiti. To lahko naredimo kar s filtriranjem:

```
df_javni = df[df['sektor']=='Javni sektor']
df_zasebni = df[df['sektor']=='Zasebni sektor']
```

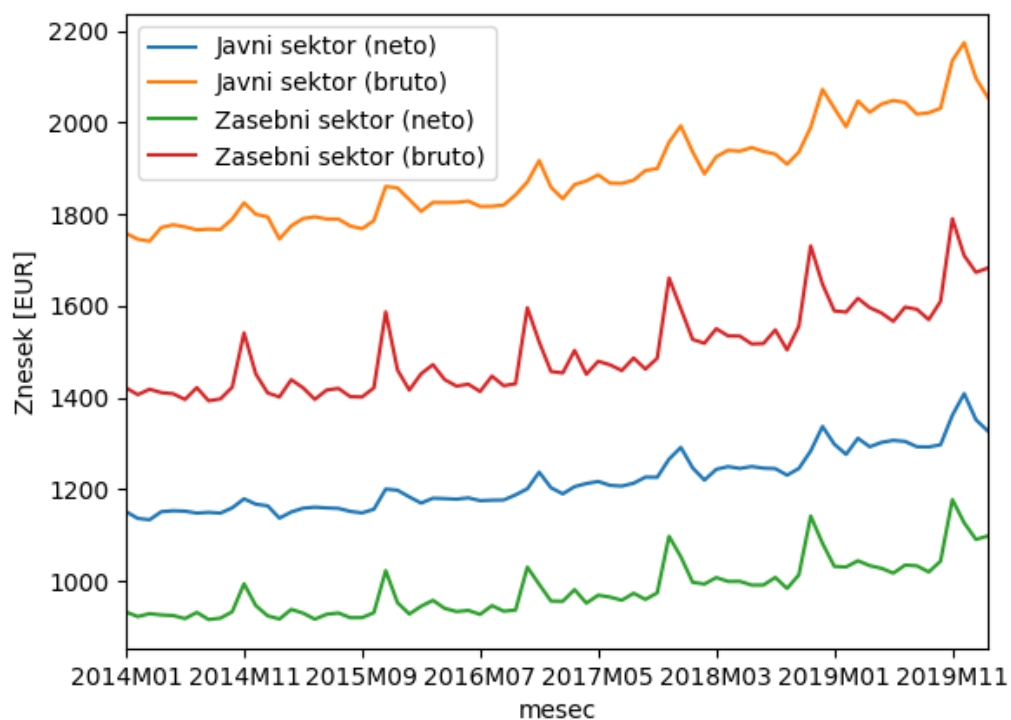
Zdaj izrišemo oba sektorja. Če ju hočemo narisati na skupnem grafu, oziroma isti osi (angl. *axis*), moramo to eksplicitno podati. Najprej pridobimo trenutno os preko funkcije `gca`:

```
ax = plt.gca()
```

Potem os podamo pri risanju, skupaj z ostalimi argumenti. Graf lahko dopolnimo še z legendo, oznakami itd.

```
df_javni.plot(x='mesec',
              y=['neto', 'bruto'],
              ax = ax)
df_zasebni.plot(x='mesec',
                y=['neto', 'bruto'],
                ax = ax)
plt.legend(['Javni sektor (neto)',
            'Javni sektor (bruto)',
            'Zasebni sektor (neto)',
            'Zasebni sektor (bruto)'])
plt.ylabel('Znesek [EUR]')
plt.show()
```

Tako dobljen graf prikazuje slika 16.2. Graf bi lahko na enostaven način spremenili



Slika 16.2 Črtni graf z ločenim izrisom za javni in zasebni sektor.

v kakšen drug tip, tako da bi nastavili izbirni argument `kind` na kakšno drugo vrednost, npr. `'bar'` za stolpčni diagram ali pa `'box'` za izris s kvartili.

Graf bi lahko narisali tudi tako, da bi podatke iz strukture `dataframe` pretvorili npr. v strukturo `ndarray` z uporabo metode `values` (metoda `values` zavrže imena stolpcev in kot matriko tipa `ndarray` vrne vsebino tabele). Do matrike bruto in neto plač javnega in zasebnega sektorja, bi lahko npr. prišli takole:

```
>>> J = df_javni[['bruto', 'neto']].values
>>> Z = df_zasebni[['bruto', 'neto']].values
```

Naenkrat lahko narišemo več grafov tudi tako, da funkciji `plot` preko vmesnika `matplotlib.pyplot` podamo kar matriko. Funkcija bo za vsak stolpec matrike izrisala svoj graf.

```
>>> plt.plot(J)
>>> plt.plot(Z)
>>> plt.show()
```

16.6 Izvoz podatkov

Podatke lahko iz tabel `dataframe` tudi enostavno izvozimo. Če bi npr. želeli podatke za javni sektor izvoziti v datoteko CSV, bi to lahko naredili z uporabo metode `to_csv`.

```
df_javni.to_csv('place_javni.csv', index=False)
```

Opcijski argument `index` smo nastavili na vrednost `False`, saj ponavadi stolpca z indeksi ne želimo izvažati (včasih pa).

Podatke bi lahko izvozili tudi v Excelovo datoteko, in sicer z metodo `to_excel`, ki deluje zelo podobno kot metoda za izvoz v datoteke CSV:

```
df_javni.to_excel('place_javni.xlsx', index=False)
```

Na podoben način lahko Excelovo datoteko tudi uvozimo. Tokrat uporabimo funkcijo `read_excel`:

```
df_javni2 = pd.read_excel('place_javni.xlsx')
```

Osnutek

17 Okolje Jupyter

17.1 Interaktivni zvezki

Okolje Jupyter predstavlja alternativo okolju IDLE, ki pa uporablja tudi nekoliko drugačen zapis programov. Programe namreč zapisujemo v tako imenovane *interaktivne zvezke* (angl. *IPython Notebooks*), s končnico `ipynb`. Preden si okolje podrobneje pogledamo, ga moramo namestiti. Spet uporabimo orodje `pip`:

```
> pip install jupyter
```

Zdaj lahko okolje jupyter zaženemo, tako da se v ukazni vrstici našega operacijskega sistema premaknemo v mapo, kjer imamo shranjene datoteke, s katerimi bomo delali, in zaženemo ukaz:

```
> jupyter notebook
```

S tem smo pognali strežnik okolja Jupyter (angl. *Jupyter server*), s katerim se povežemo preko spletnega brskalnika, ki se po izvedbi zgornjega ukaza prav tako zažene.

17.2 Celice, tipi celic in njihovo poganjanje

Najbolje je, da delovanje okolja Jupyter poskusimo kar na živem zgledu. Zgled s plačami, ki smo ga naredili v prejšnjem poglavju, je v obliki Jupyter zvezka na voljo na povezavi. Prenesimo ga na svoj računalnik in shranimo v mapo, iz katere smo pgnali Jupyter. Zdaj bi morali datoteko z imenom `place.ipynb` videti v začetnem oknu okolja Jupyter. Odprimo jo s klikom nanjo. Vidimo, da je zvezek sestavljen iz dveh tipov celic.

Prvi tip celic nudi razlago. Zapisane so v jeziku **Markdown**, ki predstavlja relativno preprost *označevalni jezik* oziroma jezik za oblikovanje besedila. Če posamezno celico dvakrat poklikamo, lahko vidimo njeno izvorno kodo. Če hočemo celico spet pretvoriti v končno obliko, jo poženemo. To lahko naredimo s kombinacijo tip `ctrl + Enter` (poženi celico) oziroma `shift + Enter` (poženi celico in skoči na naslednjo).

Drugi tip celic vsebuje kodo v jeziku Python. Te celice so označene z oznako **In**. Poganjamo jih na enak način kot celice tipa **Markdown**. Ko določeno celico

poženemo, se pod njo pojavi njena izhodna (**Out**) celica, ki prikazuje rezultat njene izvedbe. Celice si med seboj delijo imenski prostor, kar pomeni, da lahko do spremenljivk, ki smo jih definirali v posamezni celici, dostopamo tudi iz ostalih celic.

V zvezke lahko dodajamo nove celice, celice brišemo, kopiramo in spreminjamo tipe. Podrobneje v razlago okolja Jupyter ne bomo šli, saj je zelo intuitivno za uporabo, zahteva pa nekaj vaje. Tako kot zahteva vajo in trening tudi programiranje samo. Lotimo se ga...

Osnutek