```
1  import sys
2  import os
3
4  def diskretiziraj_granicnima(vrijednosti,granicne):
5      izlaz=[]
6      for vrijednost in vrijednosti:
7          izlaz.append(pronadji_granicnu(vrijednost,granicne))
8      return izlaz
9
10 def pronadji_granicnu(vrijednost,granicne):
11     for i,granicna in enumerate(granicne):
12         if vrijednost<=granicna:
13             return i+1
14     return i+2
15
16 def divide_zero(a,b):
17     if b==0:
18         return 0.
19     return a/float(b)
20
21 directory=sys.argv[1]
22 output1=[]
23 file1=open(sys.argv[1]+'.out1','w')
24 file1.write('tekst\tkorpus\ttokeni\tttr\trecenice\tduzina.recenica\tleks.
… gust\tceste.reci\tceste.leme\tduzina.recenica.2\tleks.gust.2\tceste.reci.
… 2\tceste.leme.2\n')
25 output2=[]
26 file2=open(sys.argv[1]+'.out2','w')
27 file2.write('tekst\tkorpus\treci\tduzina.reci\tinfinitivi\ttrebati\tda\
… timen.glag\tduzina.reci.2\n')
28 reference_tok=set([e.decode('utf8').strip() for e in
… open(sys.argv[2]+'.200.tok')])
29 reference_lem=set([e.decode('utf8').strip() for e in
… open(sys.argv[2]+'.200.lem')])
30
31 for file in os.listdir(directory):
32     if file.endswith('.taglem'):
33         no_sents=0
34         no_tokens=0
35         types=set()
36         no_words=0
37         word_len=0
38         no_lexical=0
39         freq_words=0
40         freq_lemmas=0
41         no_infs=0
42         no_verbs=0
43         no_nouns=0
```

biblioteke za baratanje dokumentima

definisanje funkcija koje se kasnije pozivaju

preuzmi ime direktorijuma sa komandne linije (drugi argument)

uspostavi praznu strukturu (listu) u kojoj će se čuvati izlazni podaci

kreiraj dokument za izlazne podatke

ispiši zaglavlje buduće tabele

uvezi referentne tokene iz datog dokumenta i smesti ih u strukturu tipa skup

isto za referentne leme

uspostavi mesta/strukture za buduće vrednosti varijabli

```
44          no_trebati3s=0
45          no_trebati=0
46          no_da=0
47          no_imglag=0
48          for line in open(os.path.join(directory,file)):
49              if line=='\n':    ako se linija sastoji samo od novog reda (tj. ako je prazna),
50                  no_sents+=1    dodaj 1 broju rečenica
51              else:   u suprotnom podeli liniju na segmente razdvojene tabulatorom
52                  token,lemma,tag=line.decode('utf8').strip().split('\t')
53                  types.add(token)  dodaj prvi segment (=token) listi tipova
54                  no_tokens+=1    dodaj 1 broju tokena
55                  if tag.startswith('V'):
56                      no_verbs+=1
57                      if tag in ('Vmn','Van'):
58                          no_infs+=1
59                  if tag.startswith('N'):
60                      no_nouns+=1
61                  if lemma=='trebati':
62                      no_trebati+=1
63                      if tag.endswith('3s'):
64                          no_trebati3s+=1
65                  if lemma=='da' and tag=='Cs':
66                      no_da+=1
67                  if tag!='Z' and not tag.startswith('X'):
68                      if token.lower() in reference_tok:
69                          freq_words+=1
70                      if lemma in reference_lem:
71                          freq_lemmas+=1
72                      no_words+=1
73                      word_len+=len(token)
74                      if tag[0] not in 'PSCQI' and tag[:2]!='Va' and
… tag!='Rgp':
75                          no_lexical+=1
76
… file1.write('\t'.join((directory+'.'+file,directory,str(no_tokens),str(
… divide_zero(len(types),no_tokens)),str(no_sents),str(divide_zero(
… no_tokens,no_sents)),str(divide_zero(no_lexical,no_words)),str(
… divide_zero(freq_words,no_words)),str(divide_zero(freq_lemmas,no_words)),
… str(pronadji_granicnu(divide_zero(no_tokens,no_sents),[20,30,40,50])),str
… (pronadji_granicnu(divide_zero(no_lexical,no_words),[0.5,0.6,0.7])),str(
… pronadji_granicnu(divide_zero(freq_words,no_words),[0.3,0.4,0.5])),str(
… pronadji_granicnu(divide_zero(freq_lemmas,no_words),[0.44,0.59])))+'\n')
77
… file2.write('\t'.join((directory+'.'+file,directory,str(no_words),str(
… divide_zero(word_len,no_words)),str(divide_zero(no_infs,no_verbs)),str(
… divide_zero(no_trebati3s,no_trebati)),str(divide_zero(no_da,no_words*
… 10000)),str(divide_zero(no_nouns,no_nouns+no_verbs)),str(
… pronadji_granicnu(divide_zero(word_len,no_words),[4.5,5.5])))+'\n')
```

```
78 file1.close()
79 file2.close()
80
```