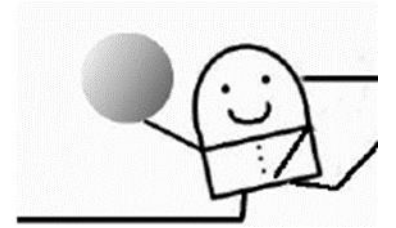


Cookbook

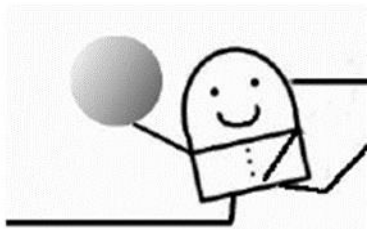
Grupo:

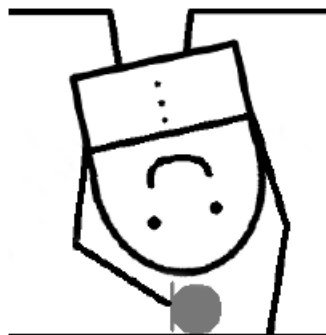
- João Lucas Yamin
- João Marcos Duda
- Fernanda Macedo
- Mariana do Vale



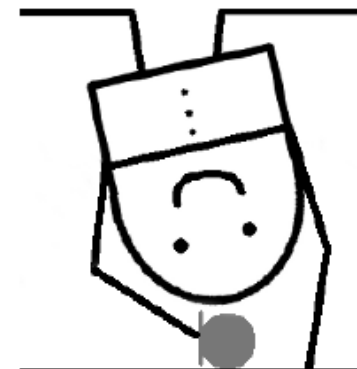


Características





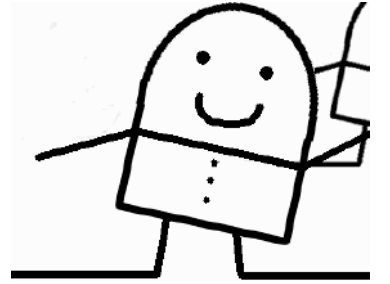
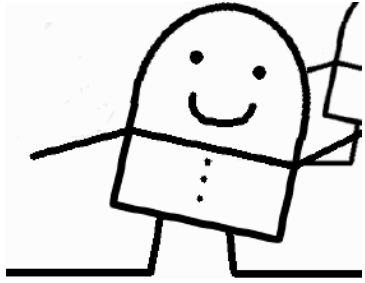
Exemplo



```
1  #!/usr/bin/env python
2  import sys, string
3
4  # The shared mutable data
5  data = []
6  words = []
7  word_freqs = []
8
9  #
10 # The procedures
11 #
12 def read_file(path_to_file):
13     """
14     Takes a path to a file and assigns the entire
15     contents of the file to the global variable data
16     """
17     global data
18     with open(path_to_file) as f:
19         data = data + list(f.read())
20
21 def filter_chars_and_normalize():
22     """
23     Replaces all nonalphanumeric chars in data with white space
24     """
25     global data
26     for i in range(len(data)):
27         if not data[i].isalnum():
28             data[i] = ' '
29         else:
30             data[i] = data[i].lower()
31
```

```
32 def scan():
33     """
34     Scans data for words, filling the global variable words
35     """
36     global data
37     global words
38     data_str = ''.join(data)
39     words = words + data_str.split()
40
41 def remove_stop_words():
42     global words
43     with open('../stop_words.txt') as f:
44         stop_words = f.read().split(',')
45     # add single-letter words
46     stop_words.extend(list(string.ascii_lowercase))
47     indexes = []
48     for i in range(len(words)):
49         if words[i] in stop_words:
50             indexes.append(i)
51     for i in reversed(indexes):
52         words.pop(i)
```

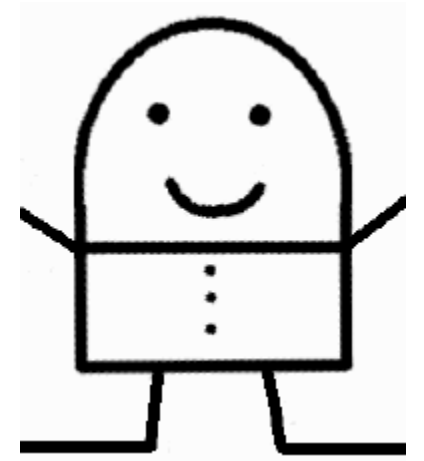
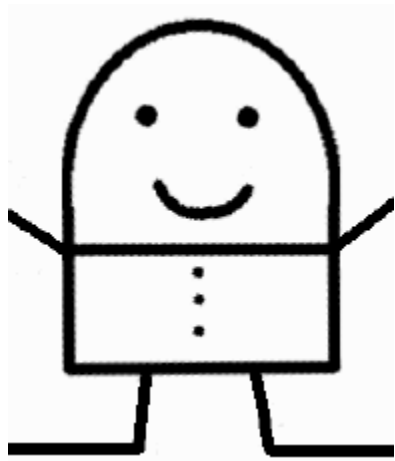
Exemplo (cont...)



```
54 def frequencies():
55     """
56     Creates a list of pairs associating
57     words with frequencies
58     """
59     global words
60     global word_freqs
61     for w in words:
62         keys = [wd[0] for wd in word_freqs]
63         if w in keys:
64             word_freqs[keys.index(w)][1] += 1
65         else:
66             word_freqs.append([w, 1])
67
68 def sort():
69     """
70     Sorts word_freqs by frequency
71     """
72     global word_freqs
73     word_freqs.sort(lambda x, y: cmp(y[1], x[1]))
```

```
76 #
77 # The main function
78 #
79 read_file(sys.argv[1])
80 filter_chars_and_normalize()
81 scan()
82 remove_stop_words()
83 frequencies()
84 sort()
85
86 for tf in word_freqs[0:25]:
87     print tf[0], ' - ', tf[1]
```

O uso do estilo CookBook





Referencia (e desenhos) :
“Exercises in Programming Style” -
Cristina Videira Lopes