Sample solutions for Exercise 3

Course: Scientific Programming (Python) by Bálint Aradi, University of Bremen

Alphabetical order

In [1]:
```
def lowered_words(txt):
    """Returns a list of lowered words in a text.

    Words a converted to lower case, unnecessary '.' and ',' characters a
    removed.

    Args:
        txt: String containing the text to process.

    Returns:
        List of words.
    """
    txt_clean = txt.replace(",", "").replace(".", "")
    words = [word.lower() for word in txt_clean.split()]
    return words
```

In [2]:
```
def print_word_list(txt):
    """Prints an alphabetical list of words occuring in a text.

    Args:
        txt: String containing the text to process.
    """
    words = list(set(lowered_words(txt)))
    words.sort()
    print("; ".join(words))
```

In [3]:
```
LOREM_IPSUM_100 = """Lorem ipsum dolor sit amet, consetetur sadipscing el
```

In [4]:
```
print_word_list(LOREM_IPSUM_100)
```

```
accusam; aliquyam; amet; at; clita; consetetur; diam; dolor; dolore; dolor
es; duo; ea; eirmod; elitr; eos; erat; est; et; gubergren; invidunt; ipsu
m; justo; kasd; labore; lorem; magna; no; nonumy; rebum; sadipscing; sanct
us; sea; sed; sit; stet; takimata; tempor; ut; vero; voluptua
```

Word occurence

In [5]:
```
def word_counts(txt):
    """Retuns a dictionary of words in a text with word counts.

    Args:
       txt: String containing the text to process.

    Returns:
        Dictionary with unique words as keys and their counts as values.
    """
    wordcounts = {}
    for word in lowered_words(txt):
        wordcounts[word] = wordcounts.get(word, 0) + 1
    return wordcounts
```

In [6]:
```
def print_sorted_word_counts(txt):
    """Prints word counts in descending order.

    Args:
       txt: Text to process.
    """
    wordcounts = word_counts(txt)
    counts = [(count, word) for word, count in wordcounts.items()]
    counts.sort(reverse=True)
    for count, word in counts:
        print(f"{word}: {count}")
```

In [7]:
```
print_sorted_word_counts(LOREM_IPSUM_100)
```

et: 8
sit: 4
sed: 4
lorem: 4
ipsum: 4
dolor: 4
diam: 4
amet: 4
voluptua: 2
vero: 2
ut: 2
tempor: 2
takimata: 2
stet: 2
sea: 2
sanctus: 2
sadipscing: 2
rebum: 2
nonumy: 2
no: 2
magna: 2
labore: 2
kasd: 2
justo: 2
invidunt: 2
gubergren: 2
est: 2
erat: 2
eos: 2
elitr: 2
eirmod: 2
ea: 2
duo: 2
dolores: 2
dolore: 2
consetetur: 2
clita: 2
at: 2
aliquyam: 2
accusam: 2

Word occurence (contracted)

In [8]:
```
def word_occurences(txt):
    """Returns word counts and words belonging to each count.

    Args:
       txt: Text to process.

    Returns:
        Dictionary with word counts as keys and list of words with that o
    """
    wc = word_counts(txt)
    occurences = {}
    for word, count in wc.items():
        # wordlist = occurences.setdefault(count, [])
        wordlist = occurences.get(count, [])
        wordlist.append(word)
        occurences[count] = wordlist
    return occurences
```

In [9]:
```
def print_word_occurences(txt):
    """Prints a sorted word occurence list.

    Args:
       txt: Text to process.
    """
    occdict = word_occurences(txt)
    occlist = list(occdict.items())
    occlist.sort(reverse=True)
    for count, words in occlist:
        print(f"{count}: {', '.join(words)}")
```

In [10]:
```
print_word_occurences(LOREM_IPSUM_100)
```

```
8: et
4: lorem, ipsum, dolor, sit, amet, sed, diam
2: consetetur, sadipscing, elitr, nonumy, eirmod, tempor, invidunt, ut, la
bore, dolore, magna, aliquyam, erat, voluptua, at, vero, eos, accusam, jus
to, duo, dolores, ea, rebum, stet, clita, kasd, gubergren, no, sea, takima
ta, sanctus, est
```

In [ ]: