Scientific Programming
(Wissenschaftliches Programmieren)

Exercise 3

Alphabetical word list

- Create a Python function / code snippet which prints all words occuring in a text in alphabetical order.
- The code should take a string as input.
- It should create a list with all the occuring words in alphabetical order and print the content of this list in one line. The words in the output should be separated by semi-colons.
- Every word should occure in the output only once, even if it occurs multiple times in the input.
- The words in the input should be converted to lower case before being processed, so that differences in upper and lower case are ignored. Also make sure to remove “.” and “,” characters from the input, so that only true words are considered.
- Test your program/function with the first 100 words of lorem ipsum. You should obtain following output:

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

Word occurance

- Create a Python function / code snippet which prints a statistics about which words and how often do occur in a text.
- The code should take a string as input.
- It should print the list of all words occuring in the input and the number of their occurance.
- It should print the most abundant worsd first and the least abundants last (ordered descendingly by occurance)
- The words in the input should be converted to lower case before being processed, so that differences in upper and lower case are ignored. Also make sure to remove “.” and “,” characters from the input, so that only true words are considered.
- Test your program/function with the first 100 words of lorem ipsum. You should obtain following output:

```
et: 8
sit: 4
sed: 4
lorem: 4
ipsum: 4
dolor: 4
diam: 4
amet: 4
voluptua: 2
…
```

Word occurance (contracted)*

- Modify the previous exercise, so that words with the same occurance are printed together.
- Test your program/function with the first 100 words of lorem ipsum. You should obtain following output:

```
8: et
4: lorem, ipsum, dolor, sit, amet, sed, diam
2: consetetur, sadipscing, elitr, nonumy, eirmod, tempor, ...
```