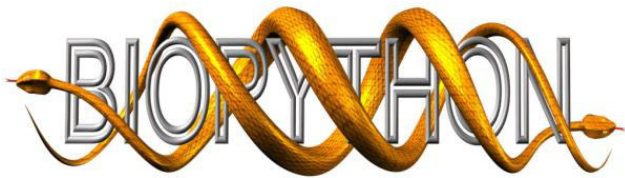
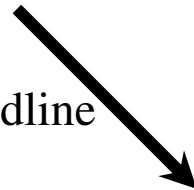


Modules

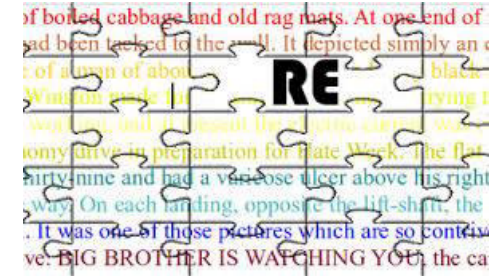
Concept of modules



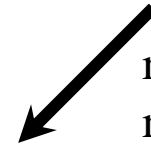
NcbiblastnCommandline
NCBIXML



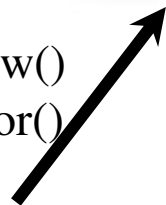
pythonTM



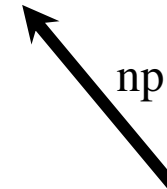
re.findall()
re.match()



plt.show()
plt.color()



np.array()



Import of modules

```
1  #basic import:
2  import re
3  #import of module under abbreviation:
4  import numpy as np
5  #import part of module:
6  from datetime import datetime
7
8  #usage of module functions:
9  re.findall()
10 datetime.now()
11
12 #try this (requires import):
13 print(str( datetime.now() ))
```

Run time calculation

- Current time is saved in two different places
- Difference is calculated to get the run time

```
1 from datetime import datetime
2
3 t1 = datetime.now()
4 #something should happen here
5 t2 = datetime.now()
6
7 print("it took " + str( t2-t1 ))
```

Regular expressions

- regular expressions (= re) enable efficient search for substrings in a given string

```
1 import re
2 some_string = "AT2G12340.1|exon-1|23745-23965|AT2G12340.2exon-1_23745-23965"
3 hits = re.findall( "AT\dG\d{5}", some_string ) #generates list of hits
4 #searches for "AT\dG\d{5}"
5 #AT, G are matching the very same character
6 #\d is matching all number 0-9
7 #{5} specifies five repetitions of the previous element
8
9 print(hits)
```

Exercises E – Part1

- 1.1) Write all AGIs of AtCol0_exons.fasta into a new file!
- 1.2) Some IDs occur multiple times. Add a filter step to reduce the results to unique IDs!
- 1.3) Calculate frequency of each AGI and construct a histogram (matplotlib)!
Tip: `plt.hist(<LIST_OF_VALUES>)`