

Biological algorithms in Python



Kevin Rue-Albrecht

Oxford Biomedical Data Science Training Programme

2020-05-05 (updated: 2020-05-04)

What is an algorithm?

- A set of rules used to solve a particular problem
- Has a specific input and output

Problem: how to make crepes?



Problem: how to make crepes?

- Inputs:
 - Ingredients
- Algorithm:
 - Preparation
 - Cooking
- Outputs:
 - Crepes

Memes

La recette de la pâte à crêpes



J'ai besoin de

Ingédients (pour environ 30 crêpes) :

- 6 œufs
- 1 pincée de sel
- 500 gr de farine
- 200 gr de sucre
- 1 litre de lait
- du beurre pour la cuisson



C'est fastoche

Préparation

Dans un grand saladier, versez la farine, le sel, le sucre, les œufs et le lait. Remuez, remuez, remuez, jusqu'à ce que vous obteniez une pâte bien lisse.



Cuisson

A l'aide d'un essuie-tout, étalez le beurre sur le fond de la poêle. Une fois que la poêle est bien chaude, versez l'équivalent d'une bonne louche de pâte. Étalez la pâte en faisant tourner la poêle. Laissez cuire un peu, puis décoller les bords et faites sauter la crêpe pour la retourner. Poursuivre la cuisson de l'autre côté.



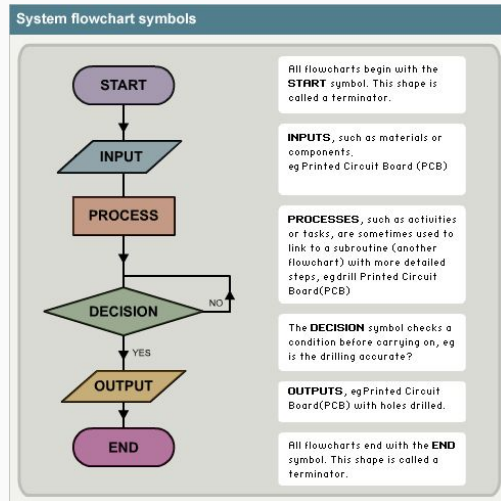
miam !

Il ne reste plus qu'à déguster !

Pseudocode

- Steps written in plain English
- Can be written as comments in the script file

Flowcharts

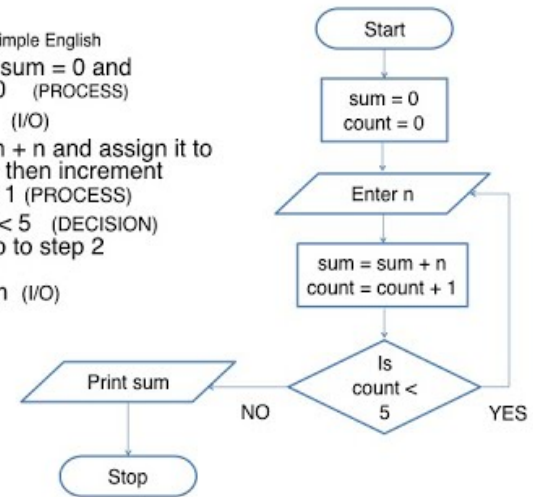


Find the sum of 5 numbers

Flowchart

Algorithm in simple English

1. Initialize $\text{sum} = 0$ and $\text{count} = 0$ (PROCESS)
2. Enter n (I/O)
3. Find $\text{sum} + n$ and assign it to sum and then increment count by 1 (PROCESS)
4. Is $\text{count} < 5$ (DECISION)
if YES go to step 2
else
Print sum (I/O)



Computer science

- Finding the maximum number in a list
- Sorting a list of numbers

5	8	6	9	1	7	3	2	4
---	---	---	---	---	---	---	---	---

Biology

- Compute GC content
- Align two sequences (e.g. DNA, protein)

Maximum Number

Find the maximum

- Input: list of numbers.

5	8	6	9	1	7	3	2	4
---	---	---	---	---	---	---	---	---

- Output: Maximum number

Exercise

- Let's write an algorithm to do that
- Start with writing pseudocode

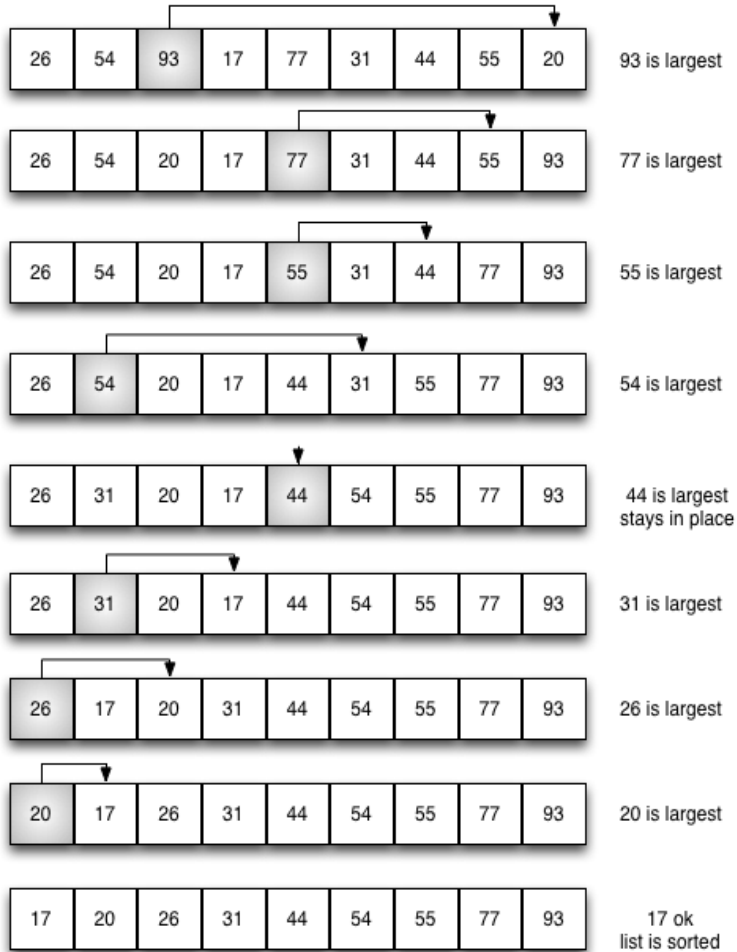
Sorting

- Input: list of numbers.

5	8	6	9	1	7	3	2	4
---	---	---	---	---	---	---	---	---

- Output: ?
- Algorithm: ?

Selection sort



Step 1

- Find the maximum across all values
- Swap it with the value at the last position.

Step 2

- Find the maximum across all values but the last one.
- Swap it with the value at the penultimate position.

Step n

- ...

Exercise

Write the selection sort in Python code.

- Use Spyder

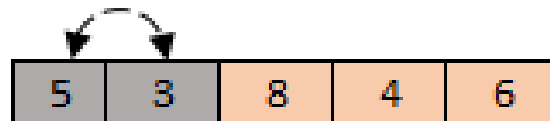
Bubble sort example

Initial



Initial Unsorted array

Step 1



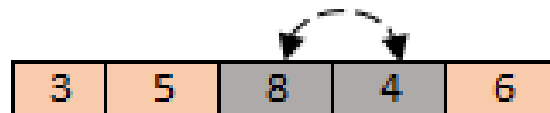
Compare 1st and 2nd
(Swap)

Step 2



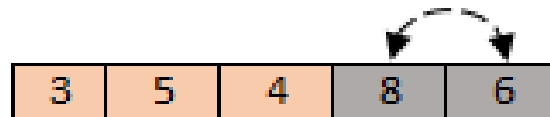
Compare 2nd and 3rd
(Do not Swap)

Step 3



Compare 3rd and 4th
(Swap)

Step 4



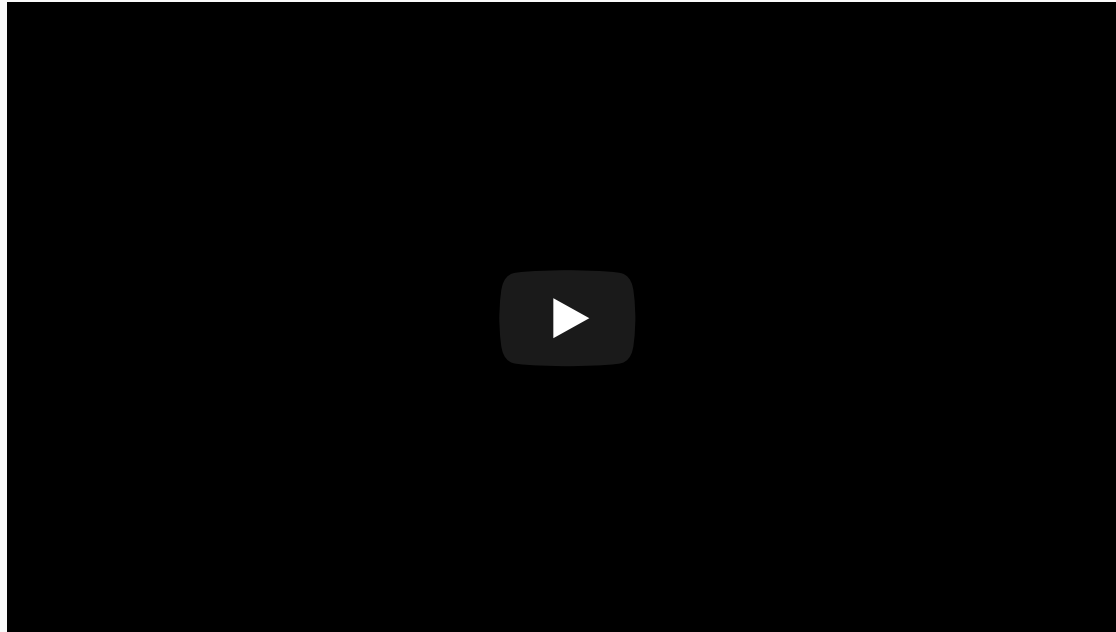
Compare 4th and 5th
(Swap)

Step 5



Repeat Step 1-5 until
no more swaps required

Bubble Sort



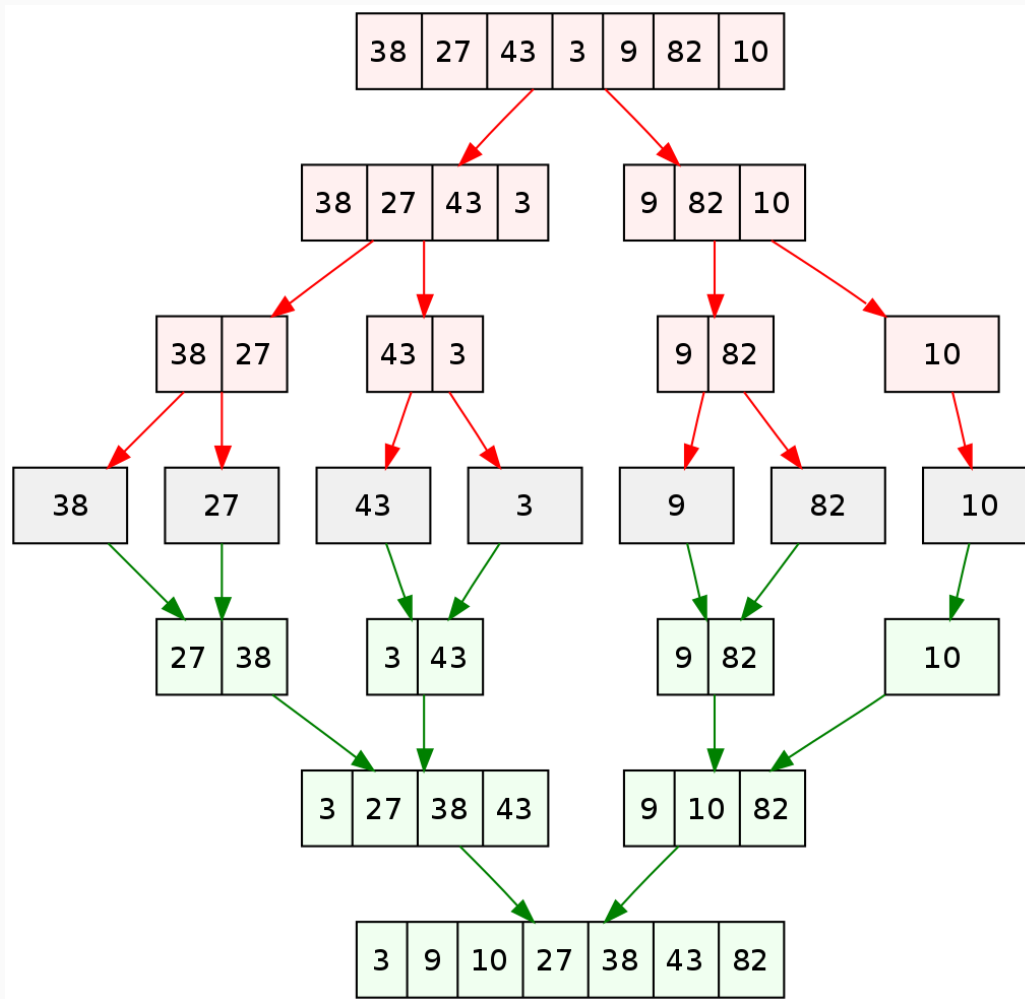
YouTube: <https://www.youtube.com/embed/nmhjrl-aW5o>

Exercise

Write the bubble sort in Python code.

- Use Spyder

Merge sort



- API: Application Programming Interface
 - <https://docs.python.org/3>
 - <https://docs.python.org/3/library/index.html>
- Find maximum: `max([6, 5, 3, 1, 8])`
- Sorting: `sorted([6, 5, 3, 1, 8])`

- Write an algorithm to compute GC content
 - <https://en.wikipedia.org/wiki/GC-content>
 - <http://rosalind.info/problems/gc>
- Test data:
 - https://www.ensembl.org/Homo_sapiens/Gene/Sequence?db=core;g=ENSG00000172216