# Ideja – Raziskovanje sortiranja podatkov

V uvodu se z dijaki pogovorimo o podatkih, ki so velikokrat (v glavnem) neurejeni in jih je treba za smiselno uporabo spraviti v urejeno obliko po nekih kriterijih. Zato morajo poznati vsaj nekaj načinov sortiranja.

1. V osnovi nato dijaki z uporabo ChatGPT poiščejo metode/načine sortiranja podatkov.

Vnos: Naštej načine oz. strategije sortiranja podatkov.

Pričakovan rezultat:

1. Mehurčkasto urejanje (Bubble sort): Primerja in zamenjuje sosednje elemente ter postopoma premika največji element na konec seznama.

2. Urejanje s vstavljanjem (Insertion sort): Postopoma vstavlja elemente na pravilno mesto v že urejenem delu seznama.

3. Urejanje s selekcijo (Selection sort): Postopoma izbira najmanjši element in ga postavlja na pravo pozicijo.

4. Urejanje z združevanjem (Merge sort): Razdeli seznam na polovico, rekurzivno ga uredi in nato združi nazaj skupaj.

5. Hitro urejanje (Quick sort): Izbere pivot, razdeli seznam na dva dela okoli pivota in rekurzivno uredi vsak del.

6. Urejanje s kopico (Heap sort): Ustvari binarno kopico iz vhodnega seznama in postopoma odstranjuje največji element.

7. Radix urejanje (Urejanje po mestih): Elemente ureja postopoma glede na posamezne števke ali mesta v njih.

1. Nato dijak opiše primer, ki ga želi dobiti za izbrano vrste sortiranja (naš primer je Bubble sort).

Sestavi program za mehurčkasto urejanje. Vhodni podatki bodo imena zapisana v datoteki "dijaki.txt" ločena z vejico. Sortirana imena zapiši v datoteko "urejeni.txt" ločeni z vejico. Kriterij sortiranja naj bo dolžina imena.

KODA:

def bubble\_sort(arr):

n = len(arr)

for i in range(n - 1):

for j in range(n - i - 1):

if len(arr[j]) > len(arr[j + 1]):

arr[j], arr[j + 1] = arr[j + 1], arr[j]

# Preberi imena iz datoteke

with open("dijaki.txt", "r") as file:

imena = file.read().split(",")

# Izvedi mehurčkasto urejanje

bubble\_sort(imena)

# Zapiši urejena imena v datoteko

with open("urejeni.txt", "w") as file:

file.write(",".join(imena))

print("Imena so uspešno urejena in zapisana v datoteko 'urejeni.txt'.")

1. Dijak naj nato pridobljeno kodo testira na podatkih, ker se pojavljajo težave.

**TEŽAVE:**

* Težava s kodiranjem – podatke se zapiše normalno s šumniki (č,š,ž)
* Datoteka shranjena v UTF-8.

Ker šumnike prevaja v dvo-znakovni zapis se dolžina imen ne izračuna pravilno.

**Naloga:**

* Odkriti težavo.
* Težavo odpraviti in popraviti kodo.
* Sestaviti primeren set podatkov za testiranje. Imena s in brez šumnikov, enake dolžine in različne dolžine.