

# Soft Computing 2023/24 - K3 - OCR

## Zadatak za grupe 1, 2, 3, 4

- Skup podataka za izradu kolokvijuma se nalazi u folderu **data1**
- Skup podataka za problem prepoznavanja ruskih reči se nalazi u folderu **data1/pictures**
- Potrebno je samostalno i programski definisati skup podataka za treniranje na osnovu dostupnih slika.
- **data1/res.csv** sadrži tačno rešenje za svaku sliku.
- Kreirati rešenje koje će ostvariti najmanji **zbir rastojanja** između tačne i prepoznate reči svih slika.
- **Rastojanje** za jednu sliku se računa na sledeći način:
  - o Ako su tačna i prepoznata reč jednake dužine, računa se **Hemingovo rastojanje**, koje predstavlja broj pozicija na kojima se dva stringa jednake dužine razlikuju:
    - $\text{rastojanje}(\text{"pera"}, \text{"pera"}) = 0$
    - $\text{rastojanje}(\text{"Pera"}, \text{"pero"}) = 2$
  - o Ako su reči različite dužine, računa se **Hemingovo rastojanje** za podstringove tačne i prepoznate reči čija je dužina jednaka dužini kraće reči i na dobijenu vrednost se dodaje apsolutna vrednost razlike dužine dve reči:
    - $\text{rastojanje}(\text{"zika"}, \text{"zika"}\text{c}) = 1$
    - $\text{rastojanje}(\text{"zika"}, \text{"ziva"}\text{c}) = 2$
- Za najveći broj bodova (**22**) potrebno je ostvariti **zbir rastojanja**  $\leq 1$ .

## Zadatak za grupe 5, 6, 7, 8

- Skup podataka za izradu kolokvijuma se nalazi u folderu **data2**
- Skup podataka za problem prepoznavanja slovačkih reči se nalazi u folderu **data2/pictures**
- Potrebno je samostalno i programski definisati skup podataka za treniranje na osnovu dostupnih slika.
- **data2/res.csv** sadrži tačno rešenje za svaku sliku.
- Kreirati rešenje koje će ostvariti najmanji **zbir rastojanja** između tačne i prepoznate reči svih slika.
- **Rastojanje** za jednu sliku se računa na sledeći način:
  - o Ako su tačna i prepoznata reč jednake dužine, računa se **Hemingovo rastojanje**, koje predstavlja broj pozicija na kojima se dva stringa jednake dužine razlikuju:
    - $\text{rastojanje}(\text{"pera"}, \text{"pera"}) = 0$
    - $\text{rastojanje}(\text{"Pera"}, \text{"pero"}) = 2$

- o Ako su reči različite dužine, računa se **Hemingovo rastojanje** za podstringove tačne i prepoznate reči čija je dužina jednaka dužini kraće reči i na dobijenu vrednost se dodaje apsolutna vrednost razlike dužine dve reči:
  - $\text{rastojanje}(\text{"zika"}, \text{"zika"}^{\text{c}}) = 1$
  - $\text{rastojanje}(\text{"zi"}^{\text{ka}}, \text{"zi"}^{\text{vac}}) = 2$
- Za najveći broj bodova (**22**) potrebno je ostvariti **zbir rastojanja**  $\leq 1$ .