

Soft Computing 2025/26 – K3 – OCR

- Skup podataka za prepoznavanje napisanih slovnih oznaka se nalazi u folderu ***data***.
- Potrebno je samostalno i programski definisati skup podataka za treniranje na osnovu dostupnih slika. Prilikom definisanja skupa podataka za treniranje zabranjeno je čuvanje novih slika na platformi. Samo treniranje modela se mora izvršavati na platformi.
- ***data/texts.csv*** sadrži tačno rešenje za svaku sliku. Prilikom izrade zadatka, koristiti vrednosti iz ovog fajla kao referentne.
- Kreirati rešenje koje će ostvariti najmanji ***zbir rastojanja*** između tačnih i prepoznatih slovnih oznaka svih slika. Pod kreiranim rešenjem se podrazumeva kod koji će izvršavati iste operacije nad svakoj od prisutnih slika (npr. nije dozvoljeno različito preprocesiranje pojedinačnih slika).
- **Rastojanje** za jednu sliku se računa na sledeći način:
 - Ako su tačna i prepoznata slovna oznaka jednake dužine, računa se **Hemingovo rastojanje**, koje predstavlja broj pozicija na kojima se dva stringa jednake dužine razlikuju:
 - $rastojanje("pera", "pera") = 0$
 - $rastojanje("Pera", "pero") = 2$
 - Ako su slovne oznake različite dužine, računa se **Hemingovo rastojanje** za podstringove tačne i prepoznate reči čija je dužina jednak dužini kraće reči i na dobijenu vrednost se dodaje apsolutna vrednost razlike dužine dve reči:
 - $rastojanje("zika", "zikać") = 1$
 - $rastojanje("zika", "živac") = 2$
- Za najveći broj bodova (**22**) potrebno je ostvariti ***zbir rastojanja < 6***.

Napomene za izradu i slanje rešenja

- Rešenje zadatka u vidu ***Python skripte*** se submit-uje putem platforme. Detaljnije uputstvo o korišćenju platforme i politika o Akademskom Integritetu se nalaze na ***Canvas*** platformi, u folderu ***Files/Vežbe/***.
- Moguće je raditi u *Jupyter Notebook* okruženju, ali se kao rešenje **mora** poslati *Python* skripta. Generisanje skripte od *Notebook*-a se vrši kroz *File* meni na sledeći način: ***File > Save and Export Notebook As > Executable Script***.
- Prilikom izvršavanja, potrebno je da skripta ispisuje konačni rezultat rešenja bez pratećeg teksta.
- **Broj osvojenih bodova se formira na osnovu postignutog rezultata i znanja prikazanog na usmenoj odbrani.**