

## RUDARJENJE IN ORGANIZACIJA BESEDIL

Udeleženci bodo uporabili metode za predprocesiranje besedil in se nato osredotočili na ekstrakcijo pomenskih delov (entitet, povezav, koreferenc). Zajete oziroma semantično označene podatke se bodo naučili shranjevati v semantičnih podatkovnih zbirkah (kot je na primer DBPedia) in po njih tudi poizvedovati. Po izobraževanju bodo udeleženci sposobni sami s pomočjo programskega jezika izvesti izbrano analizo besedila od začetka do konca.

### Namen:

Namen izobraževanja je, predstaviti in uporabiti glavne tehnike rudarjenja besedil, predvsem na semantičnem nivoju.

### Vsebina:

- Predprocesiranje in obogatitev besedil.
- Pomen jezikovnih virov, njihova izgradnja in uporaba.
- Prepoznavanje imenskih entitet, ekstrakcija povezav in odkrivanje koreferenčnosti.
- Osnovna uporaba nevronske mreže nad besedili.
- Primerjava jezikovnih lastnosti med različnimi jeziki.
- Hranjenje podatkov v obliki ontologij in iskanje po njih.
- Praktični primeri in samostojno delo za vsako izmed zgornjih tematik (Python notebook).

### Predznanje:

- osnovno predznanje iz programiranja (poljuben programski jezik)

**Trajanje:** 2x4 šolske ure

**Ključni izvajalci:** doc. dr. Slavko Žitnik, as. Ajda Pretnar