



UNIVERZITET U NOVOM SADU  
FAKULTET TEHNIČKIH NAUKA



---

# Sistemi baza podataka

---

**MongoDB –PROJEKAT**

# Obaveze

- **Prezentacija – od 29. maja do 3. juna**
- **Odbrana projekta – 15. jun**

# Prezentacija

- Na nivou tima (2 studenta ) odabrati skup podataka sa kaggle.com sajta.
  - Ili nekog drugog sajta sa reprezentativnim skupovima podataka.
  - Skup podataka bi trebalo da bude ne manji od 100MB, i da je dovoljno „širok“ kako bi se mogla vršiti konkrenta analiza (postavljanje konkretnih 10 pitanja).
- Opisati skup podataka – tema, datoteke, broj zapisa i tome slično.
- Opisati semantiku postojećih kolona – šta koja predstavlja.
- Opisati jedan primer prostiranja podataka kroz datoteke.
- Dati predlog logičke sheme baze podataka – modelovanje u dokument orijentisanim bazama podataka.
  - Prilikom realizacije projekta, biće neohodno napisati program koji će podatke prebaciti u bazu, struktuirane u skladu sa predloženom šemom.
- Dati predlog agregacije nad podacima – odgovore na kojih **desetak pitanja (5 kompleksnijih pitanja po članu tima)** ćete pokušati da date.

# Realizacija projekta

1. Inicijalna priprema skupa podataka
  - U odabranom programskom jeziku, uz oslonac na biblioteku za rad sa *MongoDB*.
  - Skup bi trebalo strukturirati u skladu sa definisanom inicijalnom logičkom shemom.
2. Analiza skupa podataka pomoću *MongoDB* upitnog jezika.
  - Odgovaranje na postavljena pitanja.
  - Upiti bi trebalo da budu adekvatne kompleksnosti.
3. Analiza performansi upita, uočavanje „uskih grla“ u izvršavanju upita.
4. Optimizacija performansi upita:
  - Restruktuiranje sheme.
  - Primena tehnika indeksiranja.
5. Prerada upita kako bi radili nad novom shemom.
6. Uporedna analiza performansi upita pre i posle optimizacije.
  - Analiza razlike u performansama, efekata postignutih optimizacijom; dijagramska predstava rezultata.
7. Vizualizacija rezultata upita uz korišćenje alata *Metabase*

# Primer projekta

- Primer projekta od prethodne godine
  - **Tema:** Analiza objava sa *Twiter*-a na temu rata u Ukrajini
  - **Autori:** Milica Vučinić i Vladimir Jovanović
- Reprezentativan primer od ranije:
  - **Tema:** Analiza popularnosti video igara pomoću podataka o prenosima sa *Twitch*-a
  - **Autori:** Milana Todorović, Ana Perišić
  - **Github repozitorijum:** <https://github.com/milana-todorovic/SBP>
  - **Tema:** Analiza skupa podataka “Formula 1 World Championship (1950-2023)”
  - **Autori:** Vuk Milanović i Andrija Stanišić
  - **Github repozitorijum:** <https://github.com/milanovicv/SBP>

**PITANJA?**