

Istraživanje podataka u bioinformatički

Bioinformatičke baze podataka

Matematički fakultet, Univerzitet u Beogradu

May 28, 2026

- **NCBI** (*National Center for Biotechnology Information*) predstavlja najvažniji bioinformatički portal koji integriše veliki broj baza podataka i alata za analizu bioloških podataka.
- Koristi se za pretragu sekvenci, naučnih radova, analizu gena i proteina, kao i za poređenje sekvenci različitih organizama.
- U okviru NCBI sistema nalaze se:
 - **GenBank** – baza nukleotidnih sekvenci DNK i RNK koja sadrži sekvence velikog broja organizama.
 - **PubMed** – baza naučnih radova i biomedicinske literature iz oblasti medicine, biologije i bioinformatike.
 - **BLAST** (Basic Local Alignment Search Tool) – alat za poređenje bioloških sekvenci i pronalaženje sličnih DNK, RNK ili proteinskih sekvenci.
 - **GEO** (*Gene Expression Omnibus*) – baza podataka ekspresije gena i funkcionalno-genomskih eksperimenata.

<https://www.ncbi.nlm.nih.gov>

- **UniProt** (*Universal Protein Resource*) predstavlja najvažniju bazu proteinskih sekvenci i funkcionalnih anotacija proteina.
- Nastao je objedinjavanjem više proteinskih baza podataka i danas predstavlja centralni resurs za analizu proteina u bioinformatiči.
- Posebno su značajne komponente:
 - Swiss-Prot – ručno anotirani podaci,
 - TrEMBL – automatski anotirani podaci.

<https://www.uniprot.org>

- Za svaki protein UniProt može sadržati:
 - naziv proteina,
 - organizam,
 - aminokiselinsku sekvencu,
 - funkciju proteina,
 - proteinske domene,
 - 3D strukturne reference,
 - enzimske funkcije,
 - biološke procese,
 - ćelijsku lokalizaciju,
 - mutacije i varijante.
- UniProt je povezan sa velikim brojem drugih baza:
 - PDB,
 - KEGG,
 - Ensembl,
 - GO,
 - OMIM.

- **Protein Data Bank (PDB)** predstavlja centralnu bazu trodimenzionalnih struktura biomolekula.
- Sadrži strukture:
 - proteina,
 - DNK,
 - RNK,
 - proteinskih kompleksa.
- Strukture su određene metodama:
 - X-ray,
 - NMR spektroskopije,
 - cryo-elektronske mikroskopije.
- Baza je veoma važna u strukturnoj bioinformatici, molekulskom modelovanju i dizajnu lekova.

<https://www.rcsb.org>

- **KEGG** (*Kyoto Encyclopedia of Genes and Genomes*) predstavlja bazu bioloških puteva i funkcionalnih mreža.
- Povezuje:
 - gene,
 - proteine,
 - metaboličke puteve,
 - bolesti,
 - lekove.
- Koristi se za funkcionalnu interpretaciju gena i analizu metaboličkih i signalnih puteva u sistenskoj biologiji.

<https://www.genome.jp/kegg>

- **Ensembl** je genomska portal za pregled i anotaciju genoma velikog broja organizama.
- Omogućava:
 - pregled hromozoma,
 - pretragu gena,
 - analizu mutacija,
 - poređenje genoma različitih vrsta.
- Veoma je značajan u genomici i komparativnoj genomici.

<https://www.ensembl.org>

- **GEO** (*Gene Expression Omnibus*) predstavlja bazu podataka ekspresije gena i funkcionalno-genomskih eksperimenata.
- GEO se koristi za analizu ekspresije gena, poređenje zdravih i obolelih tkiva i identifikaciju biomarkera.

<https://www.ncbi.nlm.nih.gov/geo>