

1. DOI dažniai PDB archyve: kiek straipsnių publikuoja po N struktūrų?
2. Wikidata: raidžių dažniai lietuviškuose tekstuose ir palyginimas su angliškais.
3. PDB archyvas: struktūros atomų skaičius, priklausomybė nuo metų, metodo, skiriamosios gebos.
4. COD archyvas: struktūros atomų skaičius, priklausomybė nuo metų, metodo, skiriamosios gebos.
5. DRUGBANK, CC (Public Domain) duomenys: struktūros atomų skaičius.
6. Project Gutenberg: dažniausiai sutinkami žodžiai, išskyrus stop-žodžius.
7. Atviros prieigos moksliniai straipsniai (tema: bioinformatika): dažniausiai sutinkami žodžiai, išskyrus stop-žodžius.
8. Atviros prieigos moksliniai straipsniai (tema: kristalografija): dažniausiai sutinkami žodžiai, išskyrus stop-žodžius.
9. Atviros prieigos moksliniai straipsniai (tema: metaloorganiniai karkasai, angl. “metal-organic frameworks”): dažniausiai sutinkami žodžiai, išskyrus stop-žodžius.
10. UNIPROT, PDB: dažniausiai pasitaikantys peptidai.
11. ChEMBL: medžiagų pasiskirstymas pagal molekulinę masę.
12. Genbank: vidutinis geno ilgis, genų ilgio pasiskirstymas
13. Genbank: kodonų pasikartojimo dažnis priklausomai nuo rūšies.
14. PDB BIRD: molekulių pasiskirstymas pagal molekulinę masę, hidrofobiškumą; kompleksų su ligandais skiriamoji geba. Ar kompleksų baltymai kažkuo reikšmingai skiriasi nuo tų, kurie kompleksuose nepastebėti?
15. PDB sekų ilgių pasiskirstymas
16. PDB a.r. ir peptidų dažniai
17. KEGG (<https://www.genome.jp/kegg/>) Kokie baltymai (fermentai) sutinkami metaboliniuose keliuose? Kiek šių struktūrų žinoma? Ar yra metabolinių kelių, kuriuose žinomos visų fermentų struktūros? Nei vieno fermento struktūra? Kiek vidutiniškai struktūrų žinoma? Duomenų nukėlimui naudokite KEGG RESTful API (<https://www.kegg.jp/kegg/rest/>).
18. :) Pasirinkite Jus dominančią bio- ar chemoinformatinę duomenų bazę (pvz. iš https://en.wikipedia.org/wiki/List_of_biological_databases, https://academic.oup.com/nar/search-results?f_TocHeadingTitle=Database+Issue)
Aprašykite: 1) kokie duomenys pateikiami duomenų bazėje 2) kaip duomenys identifikuojami; kokie naudojami unikalūs identifikatoriai

3) ar yra/kaip veikia REST API 4) į kokius klausimus galime atsakyti, naudodami šios DB duomenis?

Parašykite programą, kuri iš Jūsų pasirinktos duomenų bazės įrašų išrenka duomenis, reikalingus atsakymui į Jūsų klausimą, ir pateikia juos tolimesniam apdorojimui tinkamu formatu (CSV, XML, TAB, etc.). Formatas turi būti tinkamas Unix įrankiams ir statistinėms programoms, tokioms kaip R.