

Házi feladat - 7. hét

Határidő: 2015. március 31. 05:00

E-mail: adatbazismb2015@gmail.com

Tárgy: HF5 CSY

Formai követelmények a scripthez:

1. név, neptun kód az első sorokban legyen
2. a szkript csak a feladatszámozással ellátott megoldásokat tartalmazza (tehát a feladat szövege nem kell)
3. a megoldás csatoltan .sql vagy .txt kiterjesztésű legyen
4. a fájlnev tartalmazza a digitusos vagy adatbázisos azonosítót és a hf számát

Adatbázis: A Signalink [1] adatbázishoz tartozó séma megtalálható a wiki-n:

<https://wiki.itk.ppke.hu/twiki/bin/view/PPKE/AdatbazisKezelesMB2015/Signalink.pdf>

Az adatbázis tartalmi leírása a Signalink weboldalán található (<http://signalink.org/>).

Feladatok:

- a) Számold össze, hogy az egyes fehérjékhez hány név tartozik a name táblából! Az eredmény három oszlopot tartsd meg: A fehérje ID-je, UNIPROT_NAME-je és a nevek száma!
- b) Számold össze, hogy adatforrásonként (source tábla) hány darab interakció van! Az eredményt rendezd az interakciók száma szerint csökkenő sorrendbe! (id, db_name, interakciók száma)
- c) Készíts statisztikát a fehérjék-hez tartozó pathway-ek és topológiák számáról! Eredmény: fehérje id, fehérje uniprot name, pathway-ek száma, topológiák száma.
- d) Számold össze, hogy hány fehérje (protein) van az egyes pathway-ekben! Az eredmény tartalmazza: pathway id, pathway name, fehérjék száma.
- e) Számold össze, hogy a fehérjék hány topológiában vesznek részt! Az eredményt rendezd a topológiák száma szerint csökkenő sorrendbe! Az eredmény attribútumai: fehérje id, fehérje uniprot name, topológiák száma.
- f) Készítsd el az emberhez tartozó (a protein táblában: NCBI_TAX_ID =9606) hálózat kifokszám-eloszlását! A kifokszám azt mondja meg, hogy az egyes fehérjék hányszor szerepelnek source-ként interakcióban. Eredmény: kifokszám, fehérjék száma
- g) Add ki fehérjénként a Pathway-ek nevét vesszővel elválasztva! Az eredményt rendezd a fehérjéhez tartozó pathway-ek száma szerint csökkenő sorrendbe! A feladathoz használd a Listagg() függvényt! Az eredmény oszlopai: protein id, protein uniprot name, pathway-ek neve vesszővel elválasztva.

I. Referenciák

[1] SignaLink 2.0 - A signaling pathway resource with multi-layered regulatory networks Fazekas D*, Koltai M*, Türei D*, Módos D, Pálffy M, Dúl Z, Zsákai L, Szalay-Bekó M, Lenti K, Farkas I J, Vellai T, Csermely P, Korcsmáros T (* equal contributions) BMC Systems Biology 2013, 7:7.