

1. ZH, 2015. március 17.

Neptun-kód:

1. Készítsen el az alábbi leírás alapján egy adatbázis logikai diagrammot (az E-R modell szintjének felel meg)! Használhatja az Oracle SQL Developer, Data Modeler lehetőségeit és "Information Engineering" notációját. Más notáció elfogadható, ha külön, egyértelműen elmagyarázza). (45%)
2. Készítse el a relációs sémadiagrammot is a logikai modell alapján, az SQL developer notációját használva, grafikusán ábrázolva! Az öröklődést típusonként egy táblával fordítsa le! ("Table for each entity") A relációs modellben egyértelműen jelezze az elsődleges és a külső kulcsokat, valamint, adott esetben, a NOT NULL megszorítást (jelölésként javasolt: PK, FK, NN)! (45%).
3. Hasonlítsa össze a fájl alapú adattárolást az adatbázis alapú adattárolással (mik az előbbi nehézségei, milyen nehézségeket old meg egy adatbáziskezelő használata)! (10%)

Mikróba közösségek vizsgálatához többféle adatbázisban tárolható információra lehet szükségünk. Az adatok értelmezéséhez szükség van a biológiai mikróbák hierarchikus osztályozására. A gyökérelmet kivéve minden csomópontnak pontosan egy szülője van, egy szülőnek tetszőleges számú gyereke lehet. A gyökércsomópontnak nincsen szülője. Egy csomópontnak mindenféleképpen van egy alapértelmezett elnevezése (ez nem feltétlenül egyértelmű). Az egyes csomópontoknak az alapértelmezetten túl többféle elnevezése is lehetséges (~szinonimák). Ugyanaz a szinonima több csomóponthoz is (!) tartozhat. Az egyes elnevezéseknek van egy kategóriájuk is.

Az adatbázis tartalmazza a mikróbák ismert szekvenciáit is. A szekvenciának van egy kategóriája (milyen jellegű szekvenciáról van szó), egy forrása (honnan származik a szekvencia), hossza és egy rövid szöveges leírása. Minden egyes szekvencia pontosan egy darab organizmusból származik, feltesszük, hogy az adatbázisunkban nincsenek kimerikus (pl. kentaúr: ló+ember) szekvenciák.

Az egyes mikróbákhoz hozzárendelhetünk olyan részszekvenciákat (genomiális markereket), amelyek csak bizonyos mikróbákban találhatók meg. A markerek első típusa: unikális marker = pontosan egy darab mikróbaosztályra specifikus. A második típusú marker többféle mikróbaosztályra is lehet specifikus egyszerre. Mindkét esetben a hierarchián belül tetszőleges szintű csomópontra lehet érvényes ez a specifikusság. Mindkét típusú markernek van egy rövid szöveges leírása és egy hossza. Jellemzőjük továbbá a módszer, amellyel generáltuk. Az unikális markerekre jellemző egyrészt, hogy pontosan megadható, hogy melyik szekvenciából származnak, kezdőponttal és végponttal, másrészt egy CIGAR string, ami megadja a marker illesztését a referencia szekvenciához. A nem unikális markerek több szekvenciára is lehetnek specifikusak. Az adatbázis azt is tartalmazza, hogy mely szekvenciák mely részéből származnak (kezdőponttal és végponttal megadva), valamint hogy a reverz komplementer szálból lettek kivonva, vagy nem. Erre a típusú markerre jellemző még, hogy mennyire robosztus: ez egy 0 és 1 közötti mérőszám.

Az egyes szekvenciákhoz különböző annotációkat is hozzárendelhetünk. Az annotáció tartalmazza, hogy melyik kromoszómára vonatkozik, ki készítette, milyen szekvencia feature-ről van szó (exon, transcript, tRNA, stb.), hol kezdődik és végződik a feature, valamint hogy az annotáció melyik szála vonatkozik (eredeti, vagy annak reverz komplementer részére).

Vannak olyan mikróba közösségek (jellemzőjük: leírás), amiknek az összetételét már ismerjük: egy közösség 1 vagy több mikróbából áll-e, valamint hogy az adott mikróba milyen arányban található az adott közösségben.

Az adatbázisban levő valamennyi entitás egy saját, jelentéssel nem bíró („surrogate key”) kulcsot kapjon, melyhez egységes jelölést használjunk (ID).