

Script FeatA_1

Mittwoch, 28. September 2016 14:22

Installer la base de donnée

1. Installer anaconda avec python 2.7 pour 64x:
 - a. <https://www.continuum.io/downloads>
2. Télécharger bases de données (bacteriaVD, interactionsVD, neg_interactions_VD, phageVD):
 - a. <https://drive.switch.ch/index.php/s/XEerm5e6ItTcOQR>
3. Installer MySQL <https://doc.ubuntu-fr.org/mysql>:
 - a. `sudo apt install mysql-server`
 - b. Changer dans le script les données d'accès à la DB (user et passe)
4. Configurer MySQL
 - a. Entrer dans mysql: `mysql --local-infile=1 -u root -p --show-warnings`
 - b. Créer une nouvelle base de donnée: "CREATE DATABASE phage_bacts;"
 - c. Changer la taille maximum des paquets que mysql peut recevoir (faire "i" si ça ne fonctionne pas faire "ii" [un des deux fichiers est pratiquement vide, prendre l'autre]):
 - i. Ouvrir fichier "my.cnf" (normalement sous /etc/my.cnf)
 1. Changer permission d'accès en 777
 2. modifier:
 - a. `key_buffer_size = 128M`
 - b. `max_allowed_packet = 200M`
 3. Changer permission d'accès en 644
 - ii. Ouvrir fichier mysqld.cnf ("/etc/mysql/mysql.conf.d")
 1. Changer permission d'accès en 777
 2. modifier
 - a. `key_buffer_size = 128M`
 - b. `max_allowed_packet = 200M`
 3. Changer permission d'accès en 644
 - d. Redémarrer mysql "service mysql restart"
5. Importer les données
 - a. Phages: `mysql -u root -p phage_bacts < /home/diogo/Bureau/PhagesVD.sql`
 - b. Bactéries: `mysql -u root -p phage_bacts < /home/diogo/Bureau/bacteriaVD.sql`
 - c. Interactions: `mysql -u root -p phage_bacts < /home/diogo/Bureau/interactionsVD.sql`
 - d. Interactions négatives: `mysql -u root -p phage_bacts < /home/diogo/Bureau/neg_interactionsVD.sql`
6. Vérifications données:
 - a. Indication du nom de la BDD que tu vas utiliser: "use phage_bacts"
 - b. Insérer les commandes SQL pour vérifier que tout a bien été insérer:
 - i. Phage (1065): `"select count(*) from Phages;"`
 - ii. Bactéries (43): `"select count(*) from Bacteria;"`
 - iii. Interactions (1065): `"select count(*) from Interactions;"`
 - iv. Neg_interactions (1065): `"select count(*) from Negative_Interactions;"`

Script

1. Installer module MySQLdb
 - a. `pip install mysql-python`
2. Installer module bioPython
 - a. `pip install biopython`
3. Créer le tableau qui va garder les domaines pour chacune des protéines dans la base de données phage_bacts:
 - a. Code SQL création de création du tableau

```
DROP TABLE IF EXISTS PROTDOM;  
CREATE TABLE PROTDOM (  
  ProtDomId int NOT NULL AUTO_INCREMENT,  
  ProtId varchar(256) DEFAULT NULL,  
  DomainAcc char(16) DEFAULT NULL,  
  Cell_id int(11) DEFAULT NULL,  
  Bacteria_Cell tinyint(1) DEFAULT NULL,  
  ProtSeq mediumtext,  
  PRIMARY KEY (`ProtDomId`)  
);
```
 - b.

Visualiser bases de données:

- <http://sqlitebrowser.org/>
- Add ppa: " `sudo add-apt-repository -y ppa:linuxgndu/sqlitebrowser` "
- Update apt-get: " `sudo apt-get update` "
- Install: " `sudo apt-get install sqlitebrowser` "
- Open: " `sqlitebrowser` "

Eventuelles problèmes avec "importMySQLdb": faire ceci:

"`sudo apt-get install python-mysqldb`"



F1

DetectDo...

Si jamais tu as un problème pour accéder à la DB utilises ça: "`sudo mysql --socket=/var/run/mysqld/mysqld.sock -p`"

Script Update Domaines F1

Mittwoch, 5. Oktober 2016 15:40

J'utilisais ce script pour actualiser les protéines qui n'avaient pas pu être analysées, soit parce que le serveur était down, soit parce qu'il atteignait le time-out (valeur par défaut = 1) soit parce que je perdais la connexion à internet.

Pour ce script il n'y a pas grand chose à faire ci se n'est l'exécuter.



F1
UpdatedD...

Script detect interactions F1

Donnerstag, 6. Oktober 2016 11:45

Installer la base de données DOMINE:

- Créer la base de données: "CREATE DATABASE domine"
- Télécharger fichier SQL (domPfam, domPgmap, domTGo, domTInteract):
 - <https://drive.switch.ch/index.php/s/XEerm5e6ltTcQQR>
- Importer les données:
 - Go desigantions: mysql -u root -p domine < /home/diogo/Bureau/BD/domTGo.sql
 - PFAM: mysql -u root -p domine < /home/diogo/Bureau/BD/domPfam.sql
 - PGMAT: mysql -u root -p domine < /home/diogo/Bureau/BD/domPgmap.sql;
 - Interact: mysql -u root -p domine < /home/diogo/Bureau/BD/domTInteract.sql;
- Cérification des données:
 - GO (23487): select count(*) from GO;
 - INTERACTION (26 219): select count(*) from INTERACTION;
 - PFAM (10 696): select count(*) from PFAM;
 - PGMAT (10 264): select count(*) from PGMAT;

Créer le tableau Score_interactions dans la base de données **phage_bacts**:

Ce tableau contiendra, pour chaque IPP le score obtenu (sauf le 0 qui est calculé plus tard)

```
CREATE TABLE `Score_interactions` (  
  `Score_Inter_Id` int(11) NOT NULL AUTO_INCREMENT,  
  `ProtBactId` varchar(256) DEFAULT NULL,  
  `ProtPhageId` varchar(256) DEFAULT NULL,  
  `Positiv_Interaction` tinyint(1) DEFAULT NULL,  
  `Interaction_Id` int(11) DEFAULT NULL,  
  `Score_result` float DEFAULT NULL,  
  PRIMARY KEY (`Score_Inter_Id`)  
);
```

Ce scripte va calculer les scores des IPP entre deux organismes (interactions) les IPP ayant un score de 0 ne sont pas enregistrées mais sont calculées plus tard (scripte suivant)



F1
countScor...

Script FreqQtdScores F1

Montag, 10. Oktober 2016 13:41

Ce script va se charger de comptabiliser la fréquence de tous les scores de chaque interaction. Le score zéro correspond au nombre d'IPP n'ayant pas obtenu de scores supérieures à ceci.

Pour ce script il est nécessaire de créer la table suivante dans la base de données **phage_bact**.

```
CREATE TABLE `QtdScores` (  
  `QS_id` int(11) NOT NULL AUTO_INCREMENT,  
  `Interaction_Id` int(11) DEFAULT NULL,  
  `Positiv_Interaction` tinyint(1) DEFAULT NULL,  
  `ScoreNumber` float DEFAULT NULL,  
  `QuantityScore` int(11) DEFAULT NULL,  
  PRIMARY KEY (`QS_id`)  
)
```



F1
FreqQtdS...

Script générer fichier pickle F1

Montag, 10. Oktober 2016 14:03

Ce script se charge de créer un fichier du type "pickle" parce qu'il m'a été demandé de fournir un script capable de générer plusieurs set de données sans que celui-ci ne soit dépendant d'aucune base de donnée ou connexion internet.



F1
createGra...

Script générer les datasets F1

Montag, 10. Oktober 2016 14:46

Ce script est utilisé pour générer les datasets, il utilise le fichier pickle



F1
GenerateDS

Script séq protéomique F2

Montag, 10. Oktober 2016 14:50

Petit rappel, cette feautre ce base sur la séquence protéique de chaque organisme. Il s'agit, ici, de calculer les pourcentages d'AA, CH et poids de toutes les protéines et combiner toutes les IPP. Sachant qu'il y aura beaucoup de registre, sera ensuite appliqué de la PCA afin de réduire tout ça. J'ai vu que avec 2 PC on arrive à une variance de plus de 90% ce qui est suffisant.

Si tu veux vérifié les résultats tu peux, pour quelques exemples, faire la somme des valeurs des champs Perc_X et voir que ça va te donné très proche de 1 (arrondi) et de même pour les champs PERC_C

Avant de lancé le script il faut créer ce tableau, c'est ici que sont enregistrées les valeurs pour chacune des protéines.

```
CREATE TABLE `AciAmin_C_WEIGHT` (  
  `AACW_Id` int(11) NOT NULL AUTO_INCREMENT,  
  `idCell` int(11) DEFAULT NULL,  
  `bool_Bacteria` int(11) DEFAULT NULL,  
  `idProtein` varchar(256) DEFAULT NULL,  
  `Perc_A` float DEFAULT NULL,  
  `Perc_C` float DEFAULT NULL,  
  `Perc_D` float DEFAULT NULL,  
  `Perc_E` float DEFAULT NULL,  
  `Perc_F` float DEFAULT NULL,  
  `Perc_G` float DEFAULT NULL,  
  `Perc_H` float DEFAULT NULL,  
  `Perc_I` float DEFAULT NULL,  
  `Perc_K` float DEFAULT NULL,  
  `Perc_L` float DEFAULT NULL,  
  `Perc_M` float DEFAULT NULL,  
  `Perc_N` float DEFAULT NULL,  
  `Perc_P` float DEFAULT NULL,  
  `Perc_Q` float DEFAULT NULL,  
  `Perc_R` float DEFAULT NULL,  
  `Perc_S` float DEFAULT NULL,  
  `Perc_T` float DEFAULT NULL,  
  `Perc_V` float DEFAULT NULL,  
  `Perc_W` float DEFAULT NULL,  
  `Perc_Y` float DEFAULT NULL,  
  `Perc_X` float DEFAULT NULL,  
  `Perc_C_C` float DEFAULT NULL,  
  `Perc_C_H` float DEFAULT NULL,  
  `Perc_C_O` float DEFAULT NULL,  
  `Perc_C_N` float DEFAULT NULL,  
  `Perc_C_S` float DEFAULT NULL,  
  `Perc_Weight` float DEFAULT NULL,  
  PRIMARY KEY (`AACW_Id`)  
);
```



F2 compter
pourcent...

Script PCA

Dienstag, 11. Oktober 2016 09:00

Ce script va chercher les registres de chaque interaction dans la base le tableau AciAmin_C_WEIGHT, appliquer la PCA et enregistrer les résultats des 2 composantes principales dans un fichier, générant ainsi le set de donnée basé sur la séquence protéique



F2_PCA