

PhyloExplorer

((Phylogenetic, Tree), Explorer)

Nicolas Auberval
Nicolas Clairon
Sorel Diser

Introduction

Contexte

Bioinformatique, phylogénie, taxonomie

Cadre

Projet (TER) de 3 mois

Sommaire

1. Définition du projet
2. Analyse et programmation
3. Bilan et conclusion
4. Démonstration

1. Définition du projet

Définition du contexte

Qu'est ce que la phylogénie?

- formalise la classification et l'évolution des espèces
- représentation sous forme d'arbre phylogénétique
- repose sur la taxonomie :
étude/description des organismes du vivant
- instabilité des modèles phylogénétiques due aux avancées moléculaires et génétiques

Analyse de l'existant

le NCBI : National Center for Biotechnology Information

Le site du NCBI

- des bases de données à disposition du public
- la gestion d'arbres phylogénétiques
- une taxonomie associée très complète

La problématique

- beaucoup d'informations, interface chargée
- peu adaptée au travail personnalisé...
(sur les arbres et collection)

Les objectifs

Le besoin principal est de pouvoir travailler sur sa propre collection d'arbres.

Fonctionnalités attendues

- parcourir et visualiser une collection d'arbres aux formats d'entrées newick ou nexus
- vérifier la conformité des arbres et taxons
- effectuer des requêtes sur une collection (tri)
- filtrer une collection (garder des sous-arbres)
- produire des statistiques sur la collection (distribution)

2. Analyse et Programmation

Les solutions retenues

- Plate-forme du projet : CherryPy / Python
- Interface web : CSS, Javascript, JQuery
- NetworkX (gestion de graphes)
- Phylogelib (gestion des arbres phylogénétiques)

- Télétravail : subversion Google
- Conduite de projet : Google code
- Méthode de travail : extrem programming

Base de données NCBI

- Base de données téléchargeable sous forme de fichiers CSV via le site NCBI
- Choix dans l'information à récupérer
- Compilation des données et regroupement dans un nouveau fichier CSV (travail de preproc.py)

Structure et interaction

TaxonomyReference

Classe permettant d'interagir avec la base de données NCBI. Elle permet de:

- vérifier si un nom est valide ou pas
- savoir si un nom est un homonyme ou non
- récupérer les parents et les parents communs à plusieurs taxons
- récupérer l'arborescence du NCBI construite à partir d'une liste de taxons

Cette classe va charger le fichier CSV en mémoire.

Structure et interaction

TreeCollection

Classe représentant une collection d'arbres. Elle permet de:

- récupérer les statistiques sur la collection
- effectuer des requêtes
- effectuer le filtrage

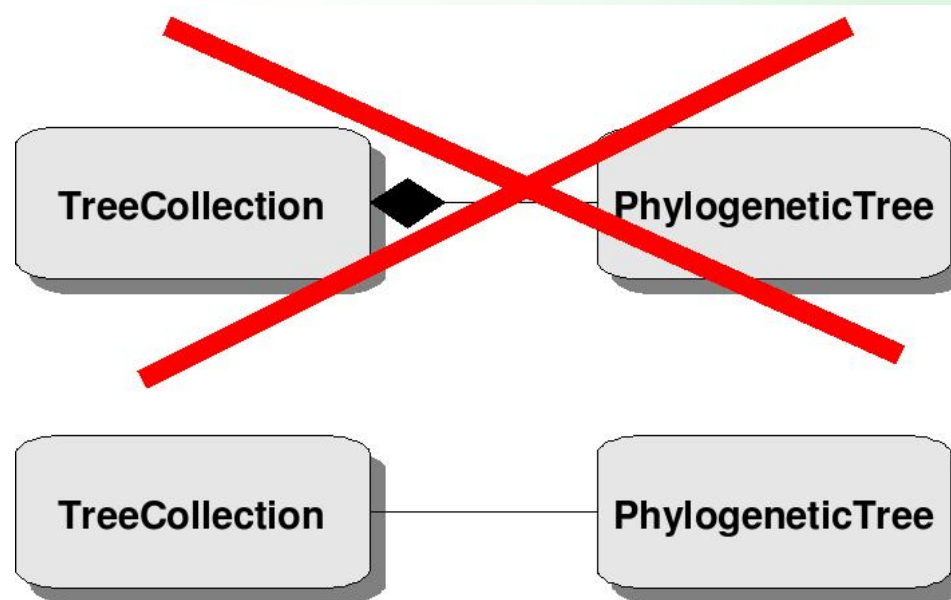
Structure et interaction

PhylogeneticTree

- Classe représentant un arbre phylogénétique.
- Elle permet d'afficher l'arbre sous forme d'arborescence.

Choix d'analyse

Rupture volontaire avec la conception objet pour gagner en rapidité :



Algorithmes délicats

- Affichage des arbres
- Requêtes
- Filtrage
- Statistiques
- Export de l'arbre NCBI au format Nexus
- Gestion des homonymes

3. Bilan et Conclusion

Cahier des charges

Accomplissement

1. affichage de l'arbre avec les noeuds internes
2. statistiques
3. dou you mean? (affichage des propositions)
4. filtrage (sous-arbre)

Cahier des charges

Valeur ajoutée

1. correction à la volée (directement dans la collection)
2. affichage des images des taxons survolés
3. téléchargement de la taxonomie du NCBI
4. téléchargement des collections filtrées
5. visualisation de l'arbre utilisateur sous forme de graphe

RÉELLE INTERACTION

Merci de votre attention ;-)

4. Démonstration