Master : STIC pour la Santé
Spécialité : « Bioinformatique, Connaissances, Données »

GMIN206 - Projet : TP XML

AGRET Clément & CHAKIACHVILI Marc

UNIVERSITÉ
DE MONTPELLIER

TIC & SANTÉ
montpellier

Présentation du projet

Le projet consiste à realiser un programme en java qui récupere les fichiers XML d'un gène orthologue d'une base de donées dans different formats depuis le NCBI. Ce programme doit être capable de manipuler ces fichiers, effectuer quelques opérations sur leurs infos contenues et créer des informations qui ne sont pas accessible, ou proposer par le site.

Notre programme à pour but de prendre en entrée le nom du fichier xml suivit d'un underscore ("_") et du format de celui-ci (BioSeq, GBCS ou Tiny(fasta)). Ex : séquence_BioSeq.

Notre application aligne les séquences quelle à récupéré grace a Chrustal W et elle renvoies la plus longue sous sequence commune au differentes séquence prise en entrée.

Réalisation

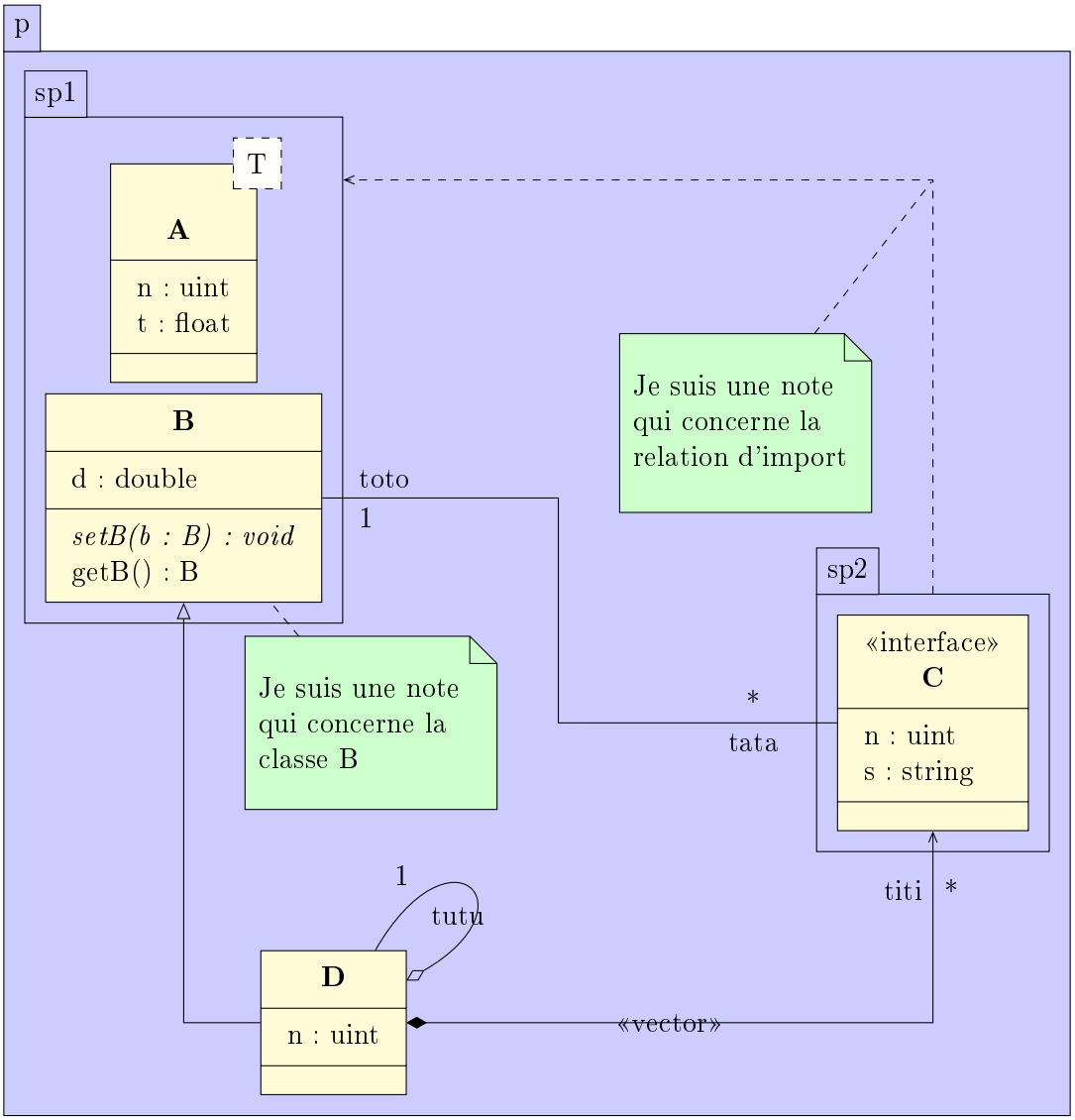
Traitement

Demande biologique : Retrouver et compter les exons présent dans les differentes séquences, retourner en sortie un fichier XML contenant uniquement les exons et leurs sous séquences associées et un alignement global des séquences de ce gène.
DTD du fichier de sortie :

Listing 1 – Fichier de sortie

```
<output>
        <gene> </gene>
        <organisms>
                <organism>
                        <GE></GE>
                </organim>
        </organisms>
        <alignment method="clustalW"> </alignment>
        <exons>
                <exon>
                        <exon_location> </exon_location>
                        <exon_sequence> </exon_sequence>
                </exon>
        </exons>
</output>
```

UML


p
sp1
T
A
n : uint
t : float
B
d : double
setB(b : B) : void
getB() : B
toto
1
Je suis une note qui concerne la relation d'import
Je suis une note qui concerne la classe B
sp2
«interface»
C
n : uint
s : string
*
tata
titi
*
1
tutu
D
n : uint
«vector»