
TP 1 : FOUILLE D'ARN

Introduction

Pour le premier tp, vous allez construire des classes permettant la manipulation de chaîne d'ARN (acide ribonucléique). Les objectifs du tp sont :

- L'utilisation de l'héritage. Plus particulièrement, vous devrez implémenter une classe qui hérite de la classe `ArrayList`.
- La construction de commentaire d'entête de méthode.

Dans les sections suivantes, nous allons décrire les interactions que votre logiciel devra offrir, les valeurs qu'il devra évaluer, les sorties qu'il va produire, les points qui serviront à évaluer votre travail et les modalités de remise de votre projet.

Interactions

Voici les différentes étapes d'interaction que votre logiciel doit suivre :

1. Demander une chaîne d'ARN.
2. Valider cette chaîne.
3. Afficher des métriques.
4. Demander une requête de recherche.
5. Valider la requête.
6. Afficher les résultats de cette requête.
7. Le programme se termine.

Aux étapes 2 et 5, si un élément est invalide, un message d'erreur descriptif est affiché et le programme se termine immédiatement.

1 : DEMANDER UNE CHAÎNE D'ARN.

Votre logiciel va demander à l'utilisateur d'entrées une chaîne d'ARN. L'utilisateur devra utiliser le clavier afin d'entrées une suite de caractères. Pour cette étape, vous **ne pouvez pas utiliser** la classe `Clavier` vu au cours inf1120. Un exemple de lecture au clavier est disponible sur le site du cours.

2 : VALIDER CETTE CHAÎNE.

La chaîne de caractères entrée par l'utilisateur doit être composée exclusivement de caractère majuscule. De plus, seuls les caractères 'A', 'C', 'G' et 'U' sont acceptés. La chaîne doit contenir au moins 6 caractères et le nombre de caractères entrés doit être un multiple de 3. Si au moins une de ces règles n'est pas vérifiée, alors un message d'erreur est affiché en utilisant l'appel de méthode suivant :

```
System.err.println( "Votre message ici." );
```

Ensuite le programme se termine en utilisant l'appel de méthode suivant :

```
System.exit( n );
```

Remplacez le 'n' par une valeur entière qui n'est pas zéro (0).

3 : AFFICHER DES MÉTRIQUES.

Ensuite, votre logiciel doit afficher des métriques décrivant l'ARN. Un ARN est composé de Nucléotides. Voici une courte description des nucléotides que nous utilisons.

Nucléotide

Les nucléotides sont des molécules de base pour la construction d'ARN. Nous allons utiliser quatre nucléotides : l'adénine (**A**), la cytosine (**C**), la guanine (**G**) et l'uracile (**U**). Ces molécules sont composées d'atomes de carbone, d'hydrogène, d'oxygène et d'azote. Elles ont aussi une masse molaire (g/mol). Le tableau suivant donne le nombre d'atomes que contient chaque nucléotide avec sa masse molaire :

Nucléotide	Nbr de carbone	Nbr d'hydrogène	Nbr d'azote	Nbr d'oxygène	Masse molaire
A	5	5	5	0	135.13
C	4	5	3	1	111.10
G	5	5	5	1	151.13
U	4	4	2	2	112.09

Votre logiciel devra calculer le nombre total de chaque type d'atome que contient l'ARN entré par l'utilisateur. Il devra aussi calculer la masse molaire totale pour l'ARN. Par exemple, l'ARN **ACGUGC** contient 27 atomes de carbone, 29 atomes d'hydrogène, 23 atomes d'azote, 6 atomes d'oxygène et à une masse molaire de 771.68 g/mol. Votre logiciel va ensuite afficher ces valeurs.

4 : DEMANDE UNE REQUÊTE DE RECHERCHE.

À cette étape, le logiciel demande une requête de recherche à l'utilisateur. La requête est lue au clavier, comme à l'étape 1. Cette requête indique ce que l'utilisateur veut trouver dans la chaîne d'ARN.

5 : VALIDER LA REQUÊTE.

Il y a deux types de requêtes possibles, le logiciel doit reconnaître le type de requête en analysant les caractères entrés :

- L'utilisateur entre un des mots suivant (la case est importante) : Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Pyl, Sec, Ser, Thr, Trp, Tyr, Val. Dans ce cas, l'utilisateur recherche un acide aminé dans l'ARN.

- b) L'utilisateur entre une suite de lettre majuscule. Dans ce cas, il doit entrer un nombre de lettres divisible par 3 et qui soit plus grand que zéro (0). Seules les lettres suivantes sont permises : A, B, C, D, G, H, K, M, N, R, S, U, V, W, Y. Dans ce cas l'utilisateur recherche une suite de nucléotides dans l'ARN.

S'il y a une erreur dans la requête, alors un message d'erreur pertinent est affiché et le programme se termine.

6 : AFFICHER LES RÉSULTATS DE CETTE REQUÊTE.

Le logiciel doit ensuite parcourir l'ARN pour trouver toutes les occurrences de l'élément demandé par l'utilisateur. Vous devez afficher la position du premier nucléotide de chaque occurrence. Le premier nucléotide d'une chaîne d'ARN est à la position 1.

- a) Pour le premier type de requête, votre logiciel doit trouver les occurrences de l'acide aminé demandé. Un acide aminé est représenté par une (1) à six (6) instances de suites de trois (3) nucléotides.

Acide aminé

Dans une chaîne de nucléotide (ARN), les éléments sont divisés en séquence de trois appelé codon. Chaque codon représente un acide aminé protéinogène. La table suivante contient les 22 acides aminés possibles avec les différentes séquences qu'ils représentent. Remarquez que certains acides aminés sont représentés par plusieurs séquences.

Abréviation	Codon(s)	Nom
Ala	GCU, GCC, GCA, GCG	Alanine
Arg	CGU, CGC, CGA, CGG, AGA, AGG	Arginine
Asn	AAU, AAC	Asparagine
Asp	GAU, GAC	Aspartate
Cys	UGU, UGC	Cystéine
Glu	GAA, GAG	Glutamate
Gln	CAA, CAG	Glutamine
Gly	GGU, GGC, GGA, GGG	Glycine
His	CAU, CAC	Histidine
Ile	AUU, AUC, AUA	Isoleucine
Leu	UUA, UUG, CUU, CUC, CUA, CUG	Leucine
Lys	AAA, AAG	Lysine
Met	AUG	Méthionine
Phe	UUU, UUC	Phénylalanine
Pro	CCU, CCC, CCA, CCG	Proline
Pyl	UAG	Pyrrolysine
Sec	UGA	Sélénocystéine
Ser	UCU, UCC, UCA, UCG, AGU, AGC	Sérine
Thr	ACU, ACC, ACA, ACG	Thréonine
Trp	UGG	Tryptophane
Tyr	UAU, UAC	Tyrosine
Val	GUU, GUC, GUA, GUG	Valine

Par exemple, l'ARN **ACGUCUUGCAGUGUU** contient l'acide aminé **Ser** aux positions 4 et 10. Il est à remarquer que les résultats possibles ne peuvent être trouvés qu'à tous les trois (3) nucléotides.

- b) Pour le second type de requête, votre logiciel doit trouver les occurrences d'une sous-séquence d'ARN. Comme pour la première requête, les résultats possibles ne peuvent être trouvés qu'à tous les trois (3) nucléotides. Certaines lettres de la requête représentent un ensemble de nucléotides. Dans ce cas, le nucléotide présent dans l'ARN doit correspondre à un des nucléotides de l'ensemble.

Codes IUPAC

L'IUPAC (Union Internationale de Chimie Pure et Appliquée) a défini des codes pour regrouper des ensembles de nucléotides. Par exemple, la lettre **R** représente un ensemble contenant une adénine et une guanine. La table suivante contient ces codes avec leur ensemble.

Code	Ensemble	Nom
R	{A, G}	<u>P</u> urine
Y	{C, U}	<u>P</u> yrimidine
S	{C, G}	<u>S</u> trong
W	{A, U}	<u>W</u> eak
K	{G, U}	<u>K</u> eto
M	{A, C}	<u>A</u> mino
B	{C, G, U}	Not A
D	{A, G, U}	Not C
H	{A, C, U}	Not G
V	{A, C, G}	Not U
N	{A, C, G, U}	<u>A</u> ny

7 : LE PROGRAMME SE TERMINE.

Le programme termine normalement.

Directives

1. Le tp est à faire seul ou en équipe de deux.
2. Votre code doit contenir une classe qui hérite de la classe `ArrayList`.
3. Commentaire :
 - a. Commentez l'entête de chaque classe et méthode. Ces commentaires doivent contenir la description de la méthode et le rôle de ces paramètres.
 - b. Une ligne contient soit un commentaire, soit du code, pas les deux.
 - c. Utilisez des noms d'identificateur significatif.
 - d. Utilisez le français.

4. Code :
 - a. Pas de `goto`, continue.
 - b. Les `break` ne peuvent apparaitre que dans les `switch`.
 - c. Un seul `return` par méthode.
5. Indentez votre code. Assurez-vous que l'indentation est faite avec des espaces.

Remise

Remettre le tp par l'entremise de Moodle. Placez vos fichiers `*.java` dans un dossier compressé de Window, vous devez remettre l'archive. Le tp est à remettre avant le 12 octobre 23 :59.

Évaluation

- Fonctionnalité (7 pts) : des tests partiels vous seront remis. Un test plus complet sera appliqué à votre tp.
- Structure (2 pt) : veillez à utiliser correctement le mécanisme d'héritage et de méthode.
- Lisibilité (2 pts) : commentaire, indentation et noms d'identificateur significatif.