

UNIVERSITÉ DE SHERBROOKE
Faculté de génie
Département de génie électrique et génie informatique

RAPPORT D'APP5

Structures de données Complexité
GIF270

Présenté à
Domingo Palao Muñoz

Présenté par
Équipe numéro 11
Benjamin Gélinas – gelb2602
Simon Leroux - lers0601

Sherbrooke – 17 mars 2021

TABLE DES MATIÈRES

1.	Développement	3
1.1	Exemple d'équation	3
1.2	Complexité de notre algorithme d'ajout de mot	3
2.	Tableaux de résultats	4

LISTE DES TABLEAUX

Tableau 1 Mots les plus souvent utilisés pour l'ensemble des auteurs	4
Tableau 2 Bi-grammes les plus souvent utilisés pour l'ensemble des auteurs	5
Tableau 3 Bi-grammes les plus souvent utilisés pour l'ensemble des auteurs	6

1. DÉVELOPPEMENT

1.1 EXEMPLE D'ÉQUATION

Les différentes structures utilisées lors de cet APP sont les listes et les dictionnaires. Nous avons choisi les dictionnaires, car celui-ci nous permettait de lier une clé (Key) et une valeur attribuée à celle-ci. Par exemple, les mots que nous avons trouvés dans un unigramme étaient liés avec leur fréquence d'apparition. En effet, le dictionnaire nous permet de rester dans un degré de complexité $\text{Big-O}(n)$, car les méthodes d'un dictionnaire ont une complexité de $\text{Big-O}(1)$ à $\text{Big-O}(n)$ dépendamment leur utilisation. Ainsi, la vitesse de calcul est rapide. De plus, la liste nous permettait de rester dans un degré de complexité assez simple qui, dans notre cas, est $\text{Big-O}(n)$ ou $\text{Big-O}(n \log n)$. Les méthodes liées aux listes avec le langage Python nous permettaient aussi d'ordonner cette liste en fonction de la fréquence.

1.2 COMPLEXITÉ DE NOTRE ALGORITHME D'AJOUT DE MOT

Dans notre algorithme pour créer les dictionnaires de mots pour les auteurs, nous avons essayé de rester dans un degré de complexité assez faible, pour limiter le temps d'exécution. Nous pouvons affirmer que notre niveau de complexité ne dépasse pas $\text{Big-O}(n)$. Nous arrivons à cette conclusion, car nous n'utilisons aucune boucle «for» imbriquée. La seule place où deux boucles pourraient avoir l'air imbriquée ne fait pas augmenter la complexité, car quand la deuxième boucle fait des itérations, la première va les sauter par la suite, ce qui fait en sorte que notre complexité n'est jamais plus grande que n . Pour ce qui est de l'algorithme de classement, nous utilisons la fonction «sorted» de python, qui elle à une complexité de $\text{Big-O}(n \log n)$.

2. TABLEAUX DE RÉSULTATS

Tableau 1 Mots les plus souvent utilisés pour l'ensemble des auteurs

Balzac		Hugo		Séguir		Verne		Voltaire		Zola	
Mots	Fréquence	Mots	Fréquence	Mots	Fréquence	Mots	Fréquence	Mots	Fréquence	Mots	Fréquence
Les	17266	Les	12403	Que	3865	Les	15453	Les	1403	Les	14158
Une	12128	Une	9253	Les	3347	Des	9437	Que	1305	Elle	12092
Que	11591	Qui	8725	Pas	3026	Que	9044	Qui	1008	Des	9457
Qui	10355	Que	8665	Elle	2632	Une	8462	Des	857	Une	9405
Des	9883	Dans	8016	Vous	2632	Qui	7062	Vous	834	Pas	6194
Elle	9659	Des	7814	Est	2472	Dans	6414	Dans	818	Que	6192
Vous	9116	Est	7650	Qui	2101	Pas	6397	Lui	684	Dans	6021
Dans	8484	Était	7494	Tour	1899	Sur	4837	Est	658	Était	5774
Pour	7688	Avait	5960	Des	1737	Est	4463	Une	609	Qui	5426
Est	7589	Pas	5726	Une	1626	Par	4408	Son	584	Avait	5138

Tableau 2 Bi-grammes les plus souvent utilisés pour l'ensemble des auteurs

Balzac		Hugo		Séguir		Verne		Voltaire		Zola	
Bigrammes	Fréquence	Bigrammes	Fréquence	Bigrammes	Fréquence	Bigrammes	Fréquence	Bigrammes	Fréquence	Bigrammes	Fréquence
Dans les	976	Jeans valjean	1108	Des ormes	393	Dans les	1008	Tous les	122	Elle avait	928
Que vous	897	Dans les	897	Que vous	383	Cyrus smith	856	Que vous	118	Elle était	757
Peut être	838	Qui est	629	Mac miche	298	Que les	743	Saint yves	118	Dans les	589
Tous les	816	Est que	590	Les enfants	248	Sur les	648	Lui dit	99	Sur les	566
Une femme	800	Tous les	573	Est pas	237	Avec une	563	Dit candide	90	Les yeux	524
Dit elle	764	Qui était	521	Est que	183	Capitaine nemo	543	Que les	78	Que les	513
Est pas	706	Toutes les	490	Tous les	175	Peut être	542	Vous avez	67	Est pas	512
Toutes les	695	Peut être	464	Elle est	171	Par les	506	Dans les	63	Les deux	443
Vous avez	690	Les deux	447	Que les	165	Que nous	428	Qui avait	60	Tandis que	435
Que les	626	Que les	445	Que nous	156	Tous les	414	Qui était	59	Tous les	404

Tableau 3 Bi-grammes les plus souvent utilisés pour l'ensemble des auteurs

	Balzac	Hugo	Ségur	Verne	Voltaire	Zola
Texte Inconnu 1	0.0213	0.0207	0.0249	0.025	0.0302	0.0104
Texte Inconnu 2	0.0237	0.0271	0.0073	0.0307	0.0269	0.0281
Texte Inconnu 3	0.0188	0.0099	0.0238	0.0212	0.0247	0.0207
Texte Inconnu 4	0.0242	0.0262	0.0257	0.0305	0.0062	0.031
Texte Inconnu 5	0.0099	0.019	0.0214	0.0212	0.0232	0.0214
Texte Inconnu 6	0.0206	0.0203	0.0272	0.0094	0.0273	0.0239