

Le problème se pose quand un client s'amène au guichet d'une banque, dit son nom et l'employé doit le retrouver dans son listing. Imaginons que le client s'appelle "Van Keersbilck". Il y a peu de chance que l'employé pense à cette orthographe et il risque donc de devoir lui demander d'épeler... L'idée serait plutôt que l'employé puisse saisir ce qu'il a entendu (par exemple venkersbilc).

L'algorithme du Soundex tente d'apporter une solution à ce problème en traduisant le nom en code 'phonétique' sur lequel la recherche dans la table sera effectuée. C'est ainsi que les mots 'mer', 'mère', 'maire', 'maisre', ... auront le même code. Cet algorithme est d'ailleurs utilisé en généalogie pour retrouver les noms qui auraient subi une transformation due, entre autres, à une faute de recopie.

L'algorithme du Soundex a été développé au début du siècle par **Margaret K. Odell** et **Robert C. Russel** au bureau américain des archives.

Voici les idées sur lesquelles s'appuie l'algorithme :

- les voyelles (dont Y) contribuent moins pour la consonance d'un mot que les consonnes. Elles seront donc supprimées sauf celle en position initiale,
- les lettres H, W ont aussi une contribution minimale et seront donc supprimées **sauf celle en position initiale**,
- les consonnes redoublées comme NN, SS et MM ou les lettres qui ont la même prononciation peuvent être réduites à une seule occurrence

Pour savoir si des lettres ont la même consonance, on s'appuie sur les tables suivantes :

pour l'anglais:		pour le français:	
Lettres	Chiffre associé	Lettres	Chiffre associé
B, F, P, V	1	B, P	1
C, G, J, K, Q, S, X, Z	2	C, K, Q	2
D, T	3	D, T	3
L	4	L	4
M, N	5	M, N	5
R	6	R	6
		G, J	7
		S, X, Z	8
		F, V	9

Voici les différentes étapes de l'algorithme :

- supprimer les éventuels 'espaces' initiaux,
- mettre le mot en majuscule,
- garder la première lettre,
- supprimer les lettres A, E, I, O, U, Y, H et W,
- remplacer les lettres restantes par le chiffre associé dans la table,
- supprimer les chiffres répétés (garder une seule occurrence),
- On ne garde que les 4 caractères de gauche du code obtenu (compléter par des espaces si nécessaire).

### Mission :

- Ecrire en java la fonction soundex(chaine, langue) qui renvoie le code de 4 caractères généré (langue = 'a' pour anglais, 'f' pour français),
- Dans un programme de test remplir une collection de String avec une dizaine de noms (aussi bien à consonance anglaise que française).
- Demander de saisir un nom et afficher les noms trouvés correspondants (essayer avec des noms anglais et français).

Par exemple :

Da Costa Tork

- |  |                    |
|--|--------------------|
| • supprimer les éventuels 'espaces' initiaux,  | <b>DaCostaTork</b> |
| • mettre le mot en majuscule,  | <b>DACOSTATORK</b> |
| • garder la première lettre,   | <b>D</b>           |
| • supprimer les lettres A, E, I, O, U, Y, H et W,  | <b>CSTTRK</b>      |
| • remplacer les lettres restantes par le chiffre associé dans la table,                                | <b>D283362</b>     |
| • supprimer les chiffres répétés (garder une seule occurrence),  | <b>D28362</b>      |
| • On ne garde que les 4 caractères de gauche du code obtenu (compléter par des espaces si nécessaire). | <b>D283</b>        |

Le code à retourner est donc D283