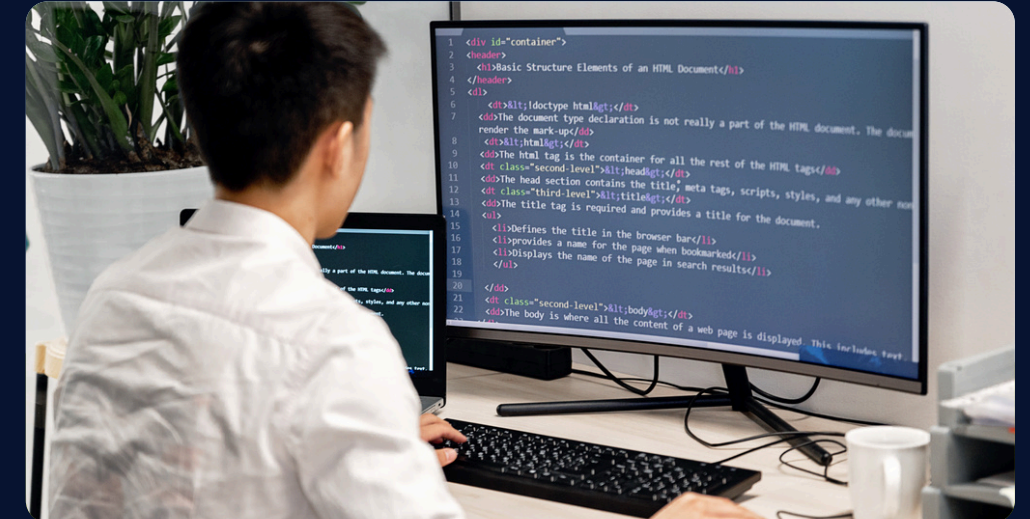


Moteur de Matching



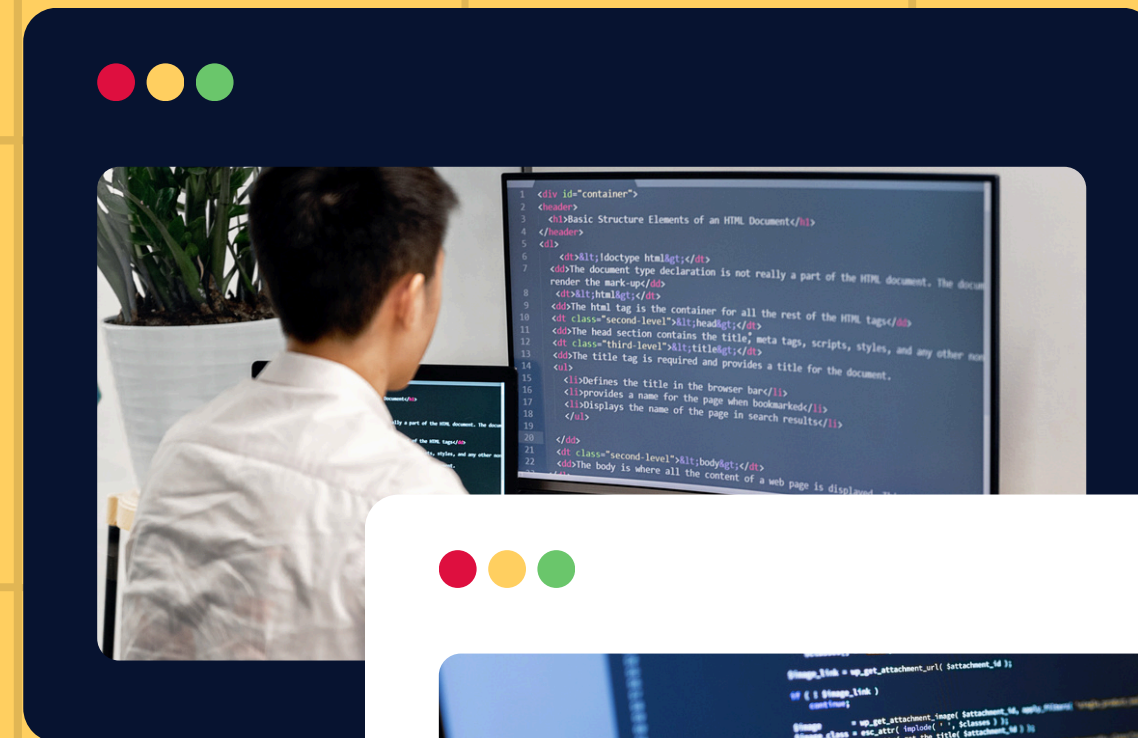
Presented By

Bsissa Adil Zanfou

Objectif du projet

Effectuer trois types de traitements sur des fichiers CSV à l'aide d'un moteur de matching basé sur des scores de similarité.

1. Recherche d'un nom dans une liste.
2. Comparaison de deux listes.
3. Déduplication d'une liste.



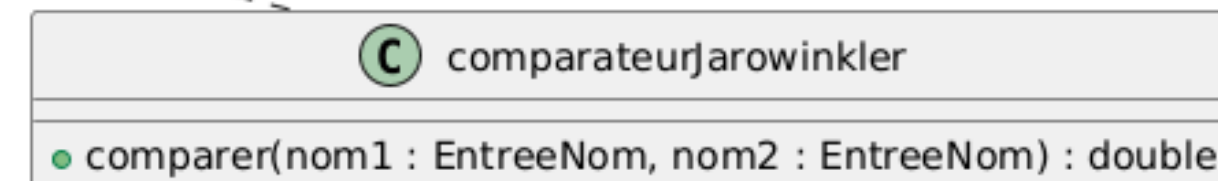
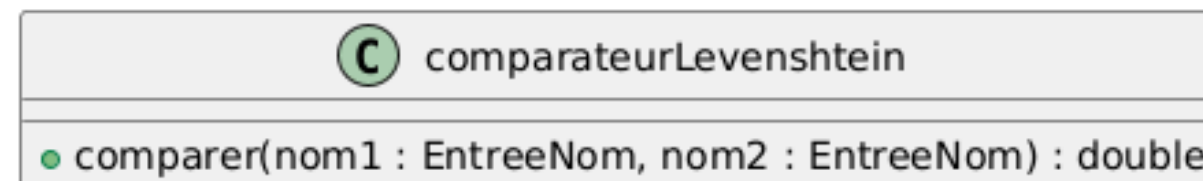
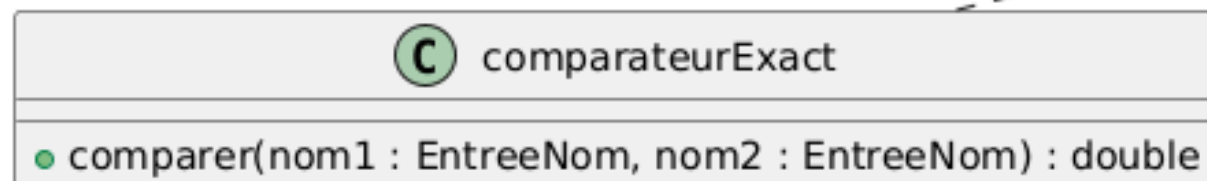
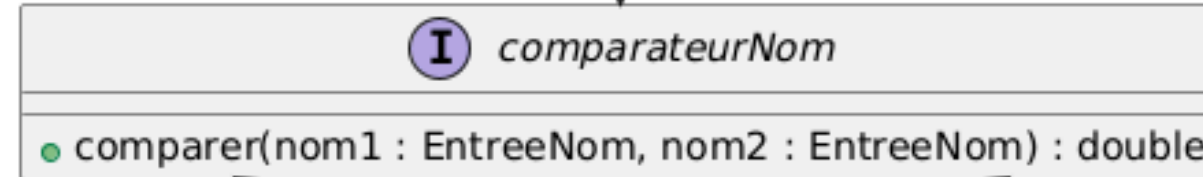
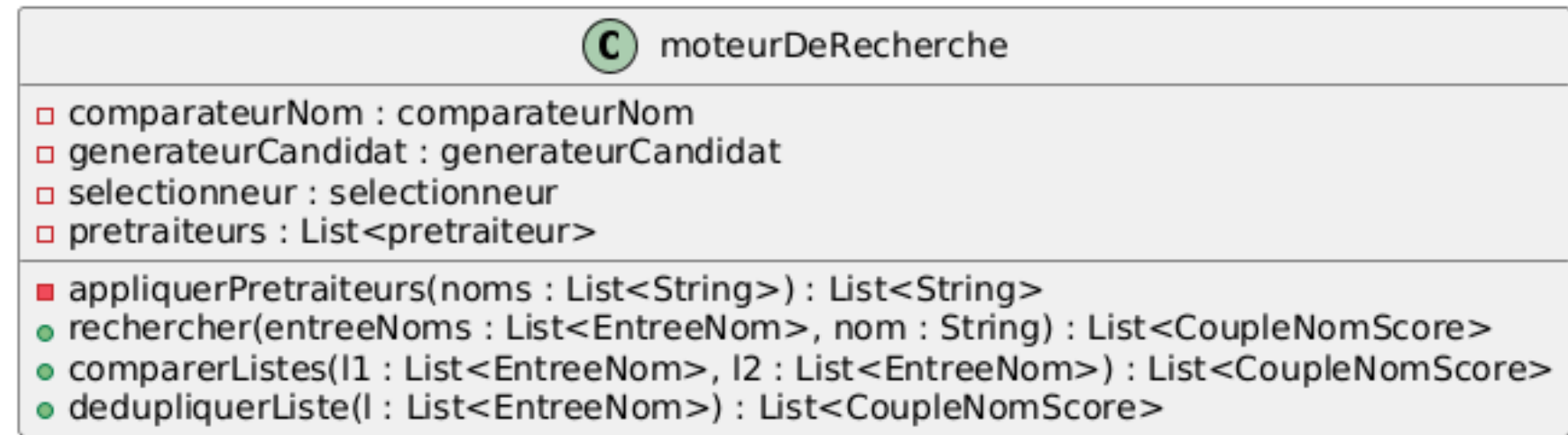
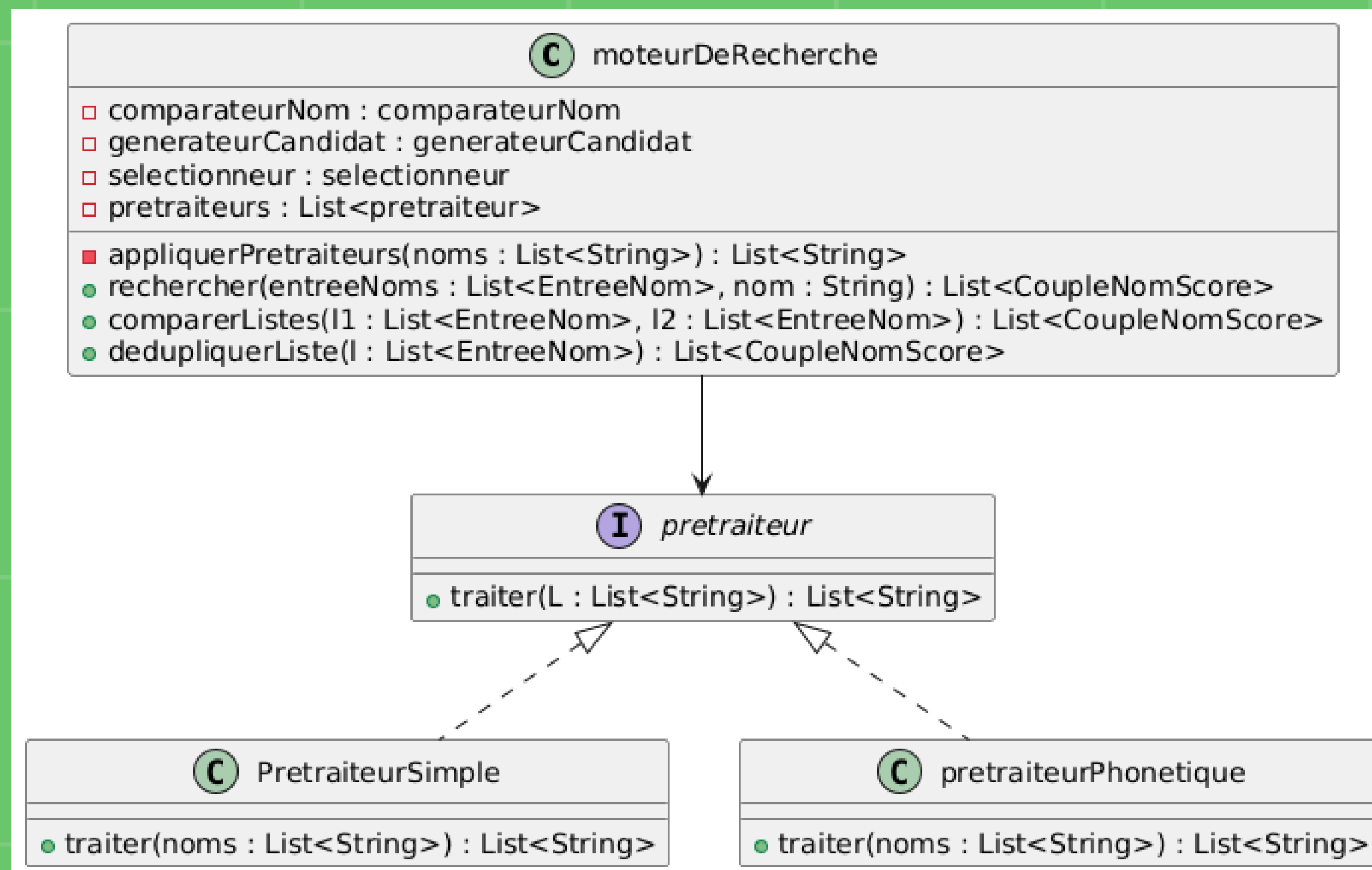
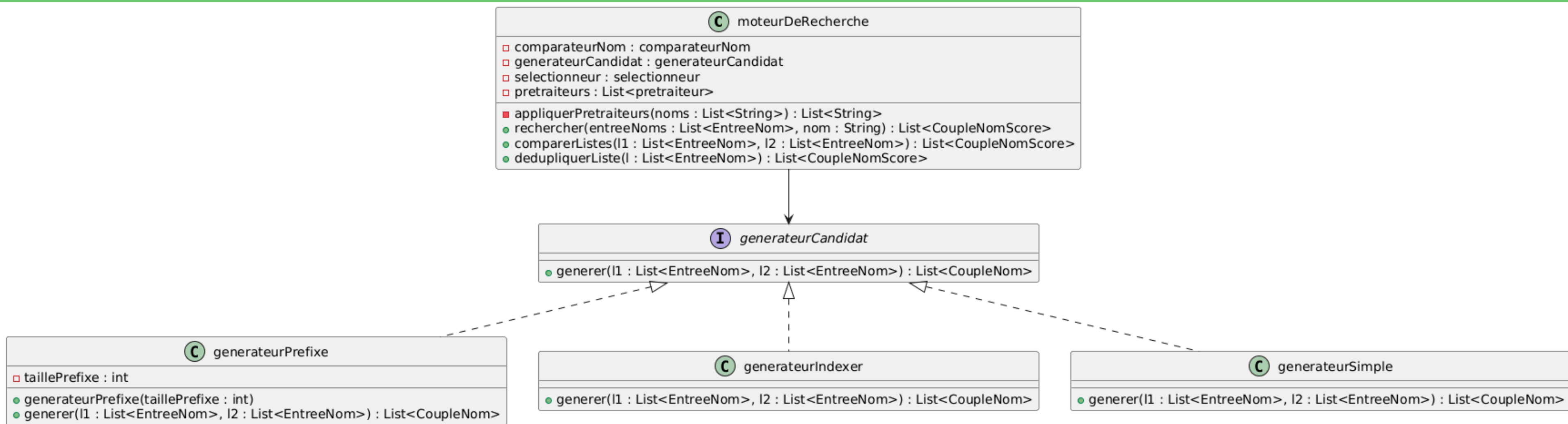
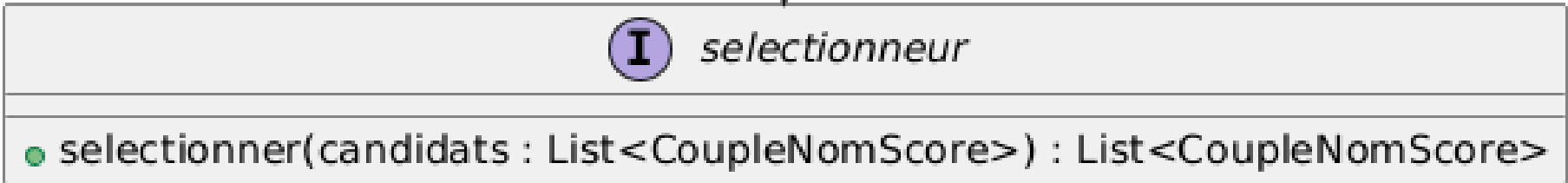
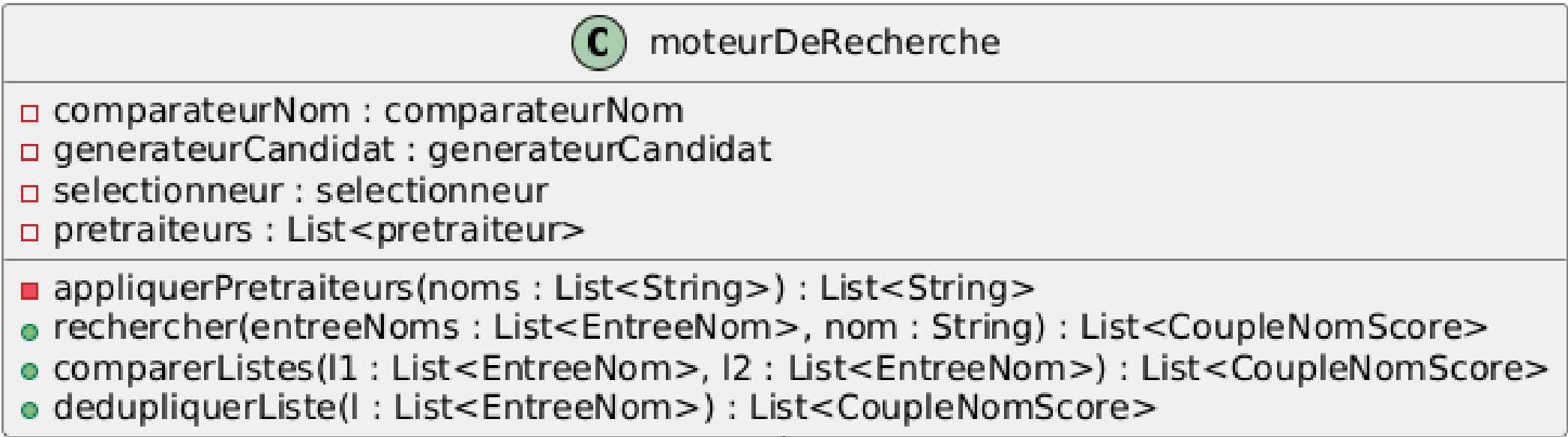
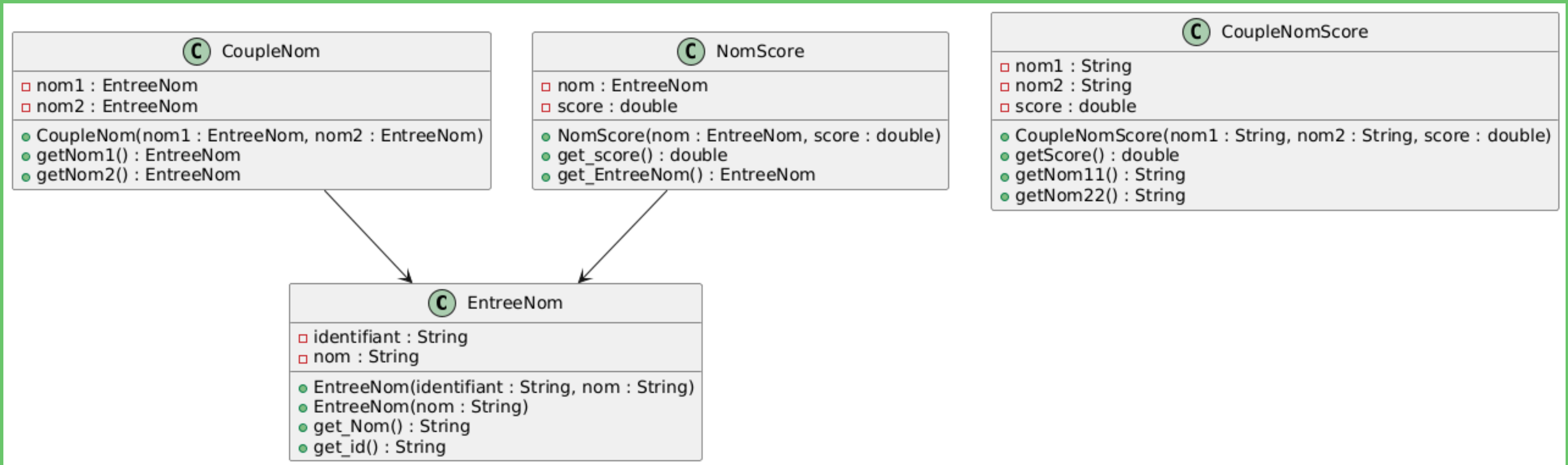


Diagramme de
classe









C indexeur

• getIndexMap(noms : List<EntreeNom>) : Map<Integer, List<EntreeNom>>

I *recuperateur*

• recuperer() : List<EntreeNom>

C recuperateurCsv

▢ fichierCsv : String

• recuperateurCsv(fichierCsv : String)

• recuperer() : List<EntreeNom>



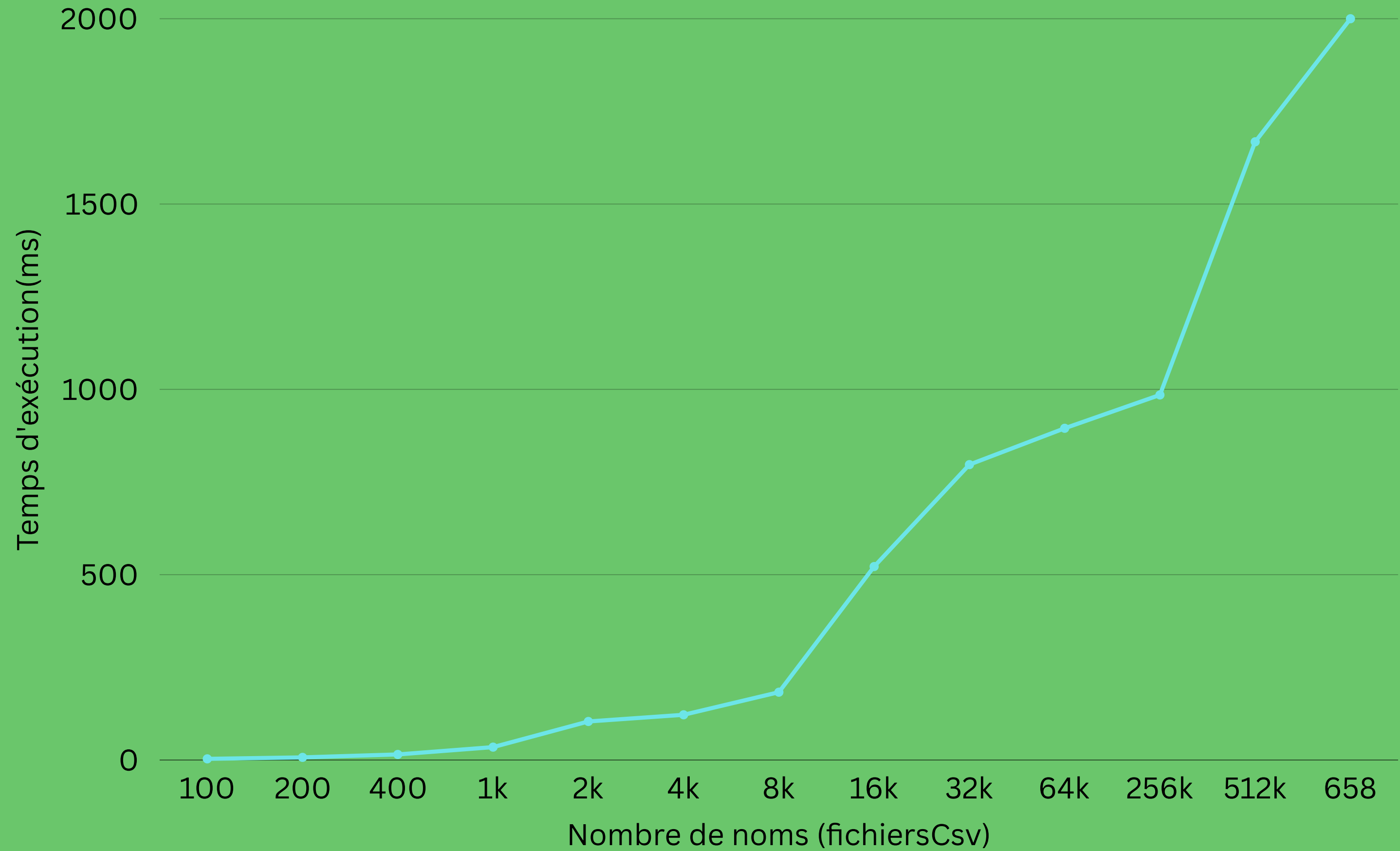
Test de la méthode rechercher

```
List<CoupleNomScore> rechercher(List<EntreeNom> entreeNoms,String nom)
```

```
Recherche de "samir" dans le fichier : C:/Users/lassa/Downloads/peps_names_400.csv  
Name: Prum Sokha, Score: 0.16666666666666663  
Name: Kim Su-gil, Score: 0.16666666666666663  
Name: DEMIDOVICH VASILIJ, Score: 0.0  
Name: Marica Montemaggi, Score: 0.0  
Name: PHIRAPHAN Salirattawiphak, Score: 0.0  
Temps d'exécution : 5.4599 millisecondes
```



- generateurSimple
- comparateurLevenschtein
- selectionneurNpremiers

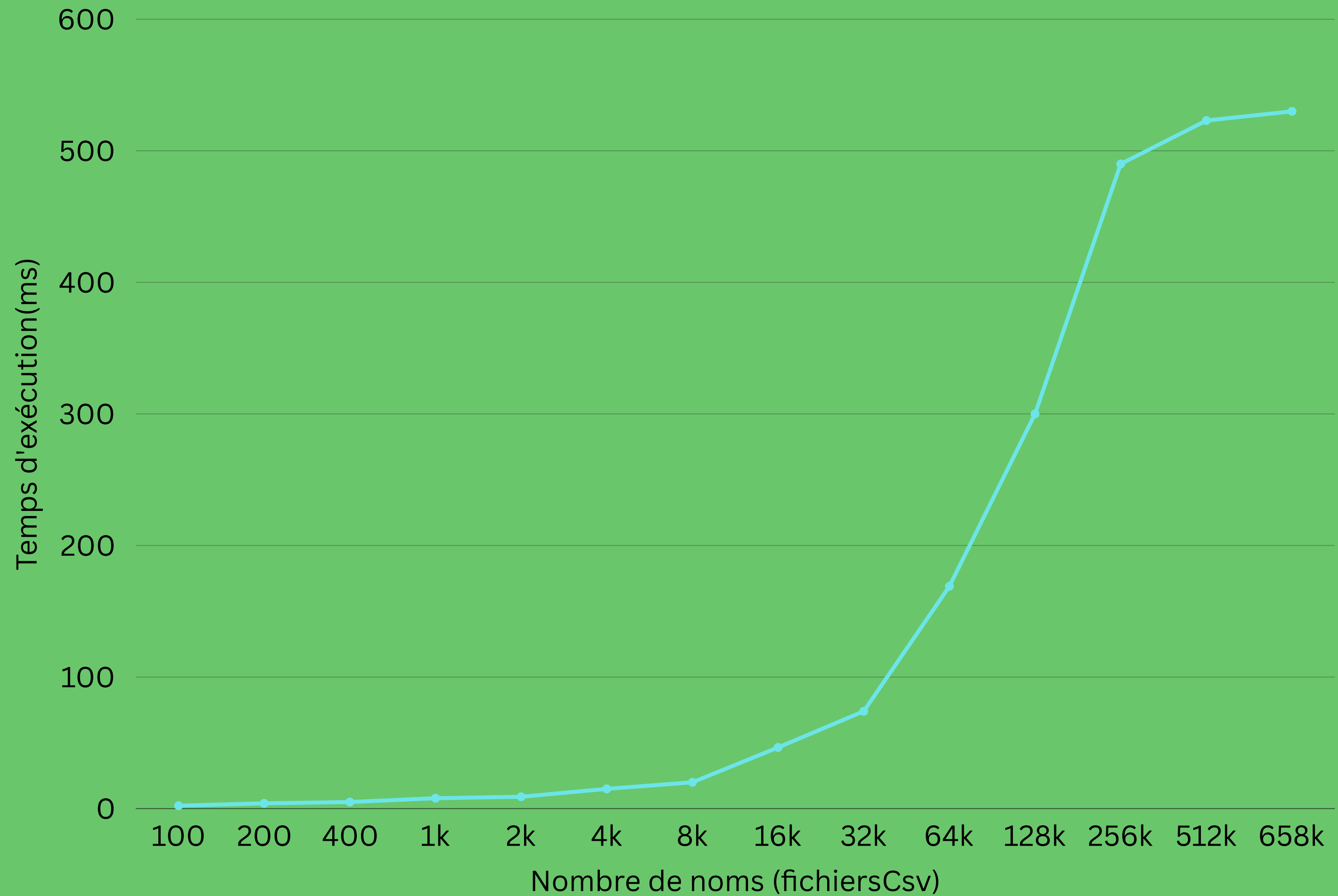


Test de la méthode rechercher


```
List<CoupleNomScore> rechercher(List<EntreeNom> entreeNoms,String nom)
```



- generateurIndexer
- comparateurLevenshtein
- selectionneurNpremiers



Test de la méthode comparer




- generateurSimple
- comparateurLevenshtein
- selectionneurNpremiers

```
public List<CoupleNomScore> comparerListes(List<EntreeNom> l1, List<EntreeNom> l2) {
```

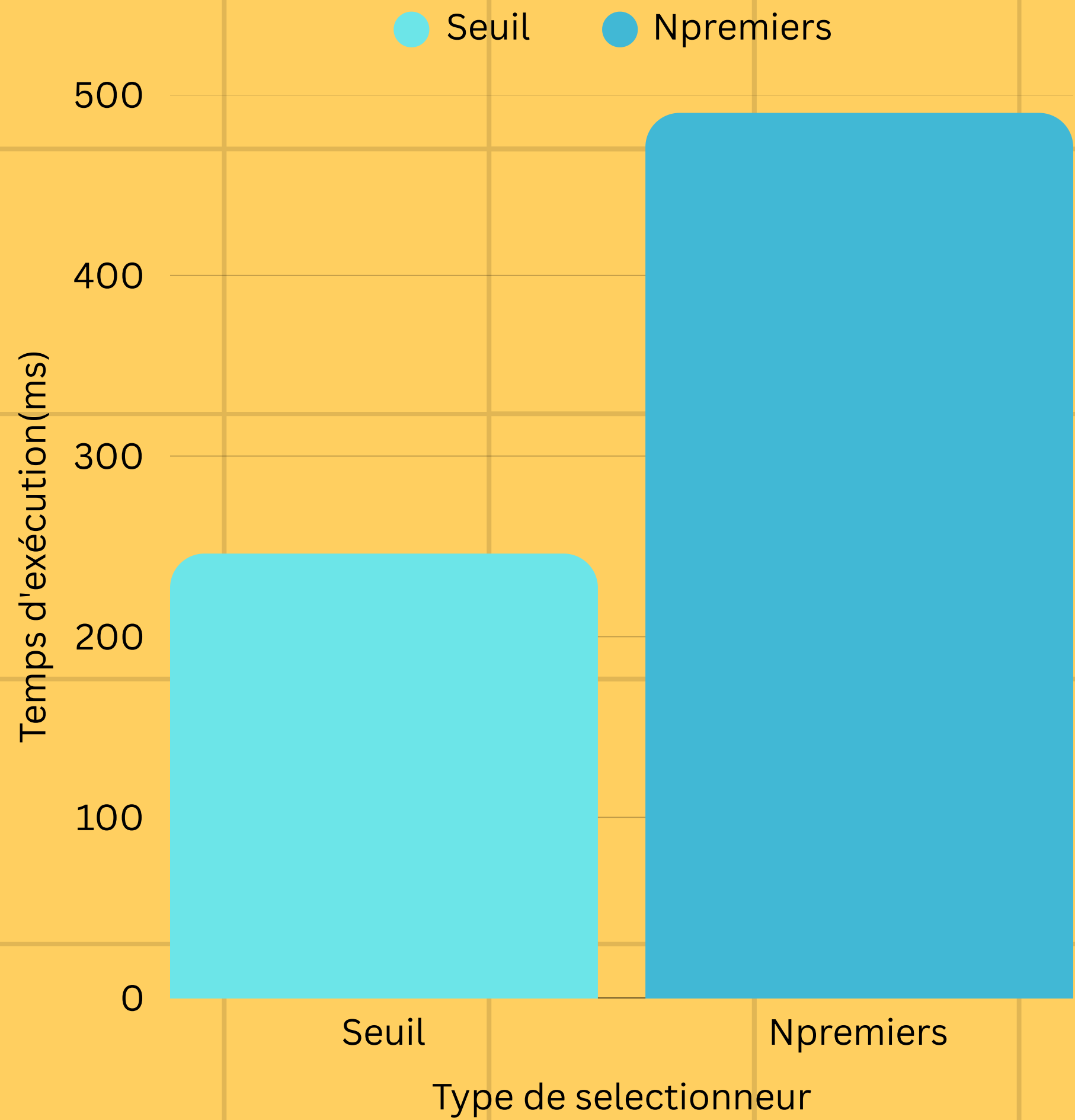
```
Nom1: Srikanta Mahata  Nom2: Souraya Moustapha Mahamat, Score: 0.5  
Nom1: Mahamady Moussa Mahamady  Nom2: Souraya Moustapha Mahamat, Score: 0.47619047619047616  
Nom1: Shah Mahatab Ahmad  Nom2: Souraya Moustapha Mahamat, Score: 0.44999999999999996  
Nom1: Issahaku Mahama  Nom2: Souraya Moustapha Mahamat, Score: 0.44999999999999996  
Nom1: Ahmat Tahir Ahmat  Nom2: Souraya Moustapha Mahamat, Score: 0.44999999999999996  
Nom1: Taban Sharifah Aate  Nom2: Souraya Moustapha Mahamat, Score: 0.44999999999999996  
Comparaison de "C:/Users/lassa/Downloads/peps_names_16k.csv" avec le fichier : C:/Users/lassa/Downloads/peps_names_8k.csv  
Temps d'exécution : 246106.1531 millisecondes
```

Test de la méthode comparer



```
-generateurSimple  
-comparateurLevenshtein  
-selectionneurSeuil
```

```
public List<CoupleNomScore> comparerListes(List<EntreeNom> l1, List<EntreeNom> l2) {
```

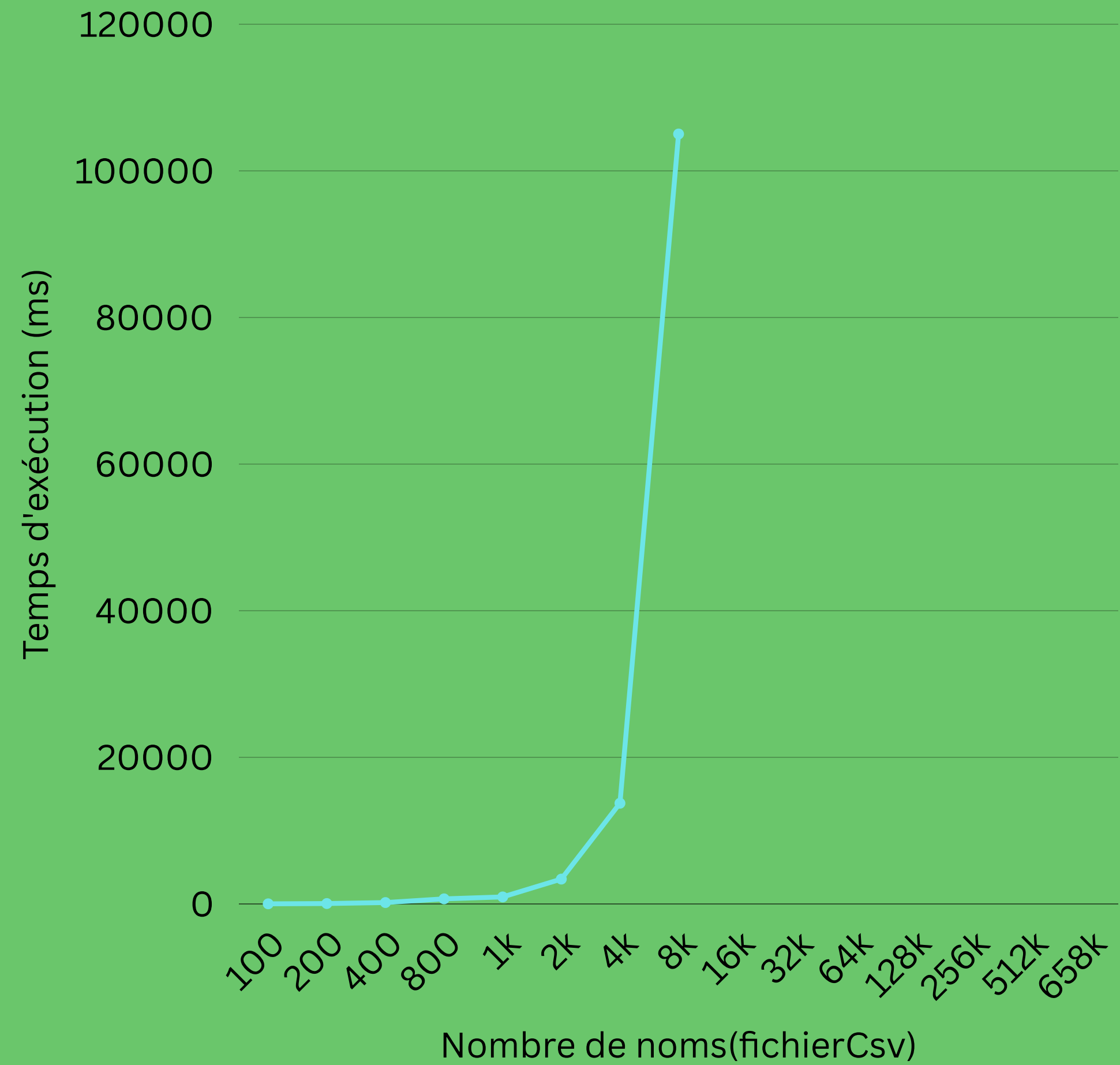


Test de la méthode dedupliquerListe

```
public List<CoupleNomScore> dedupliquerListe(List<EntreeNom> l){
```



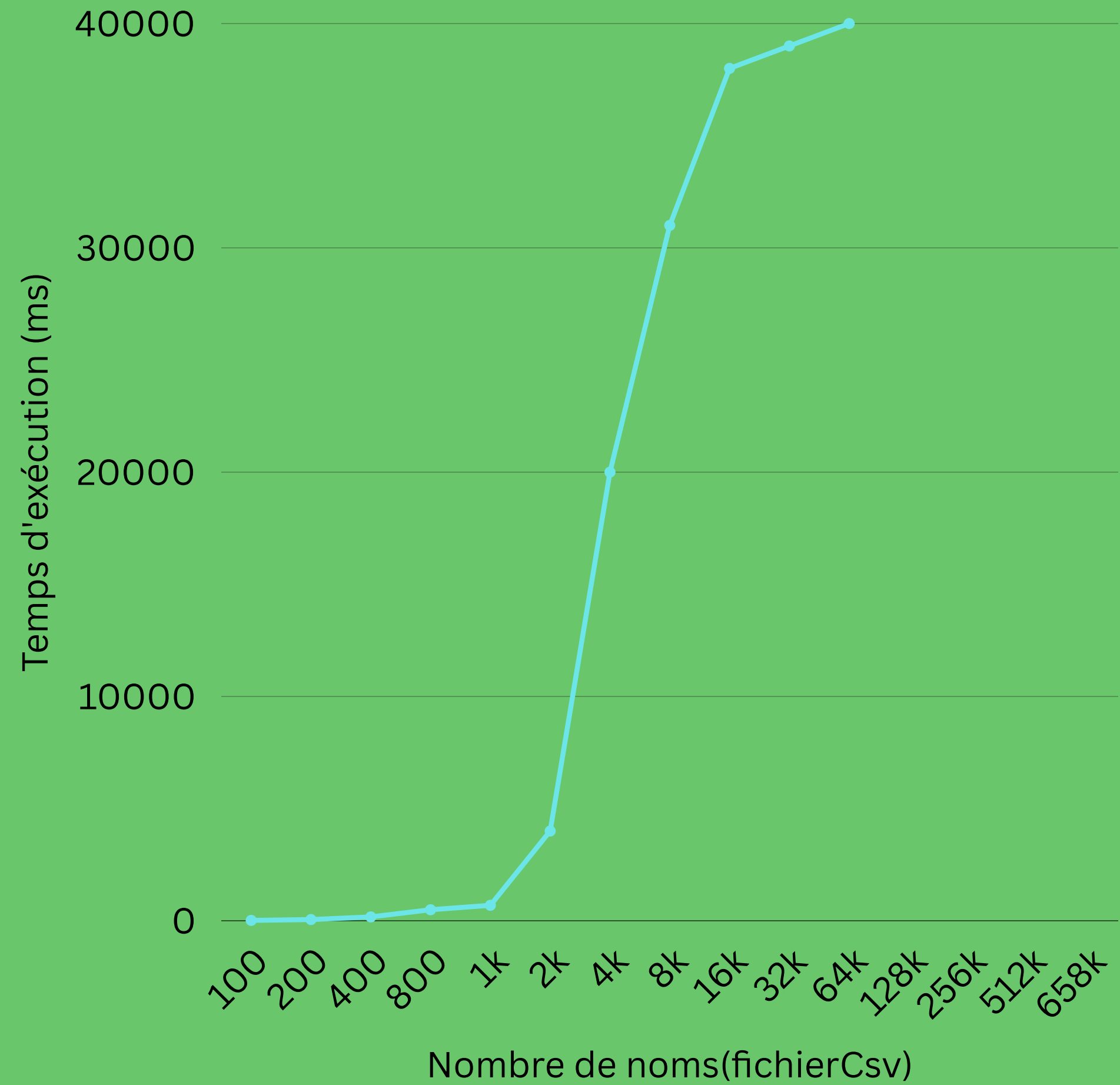
- generateurSimple
- comparateurLevenschtein
- selectionneurSeuil



Test de la méthode dedupliquerListe



```
-generateurSimple  
-comparateurJarowinkler  
-selectionneurSeuil
```





Merci pour
votre attention