

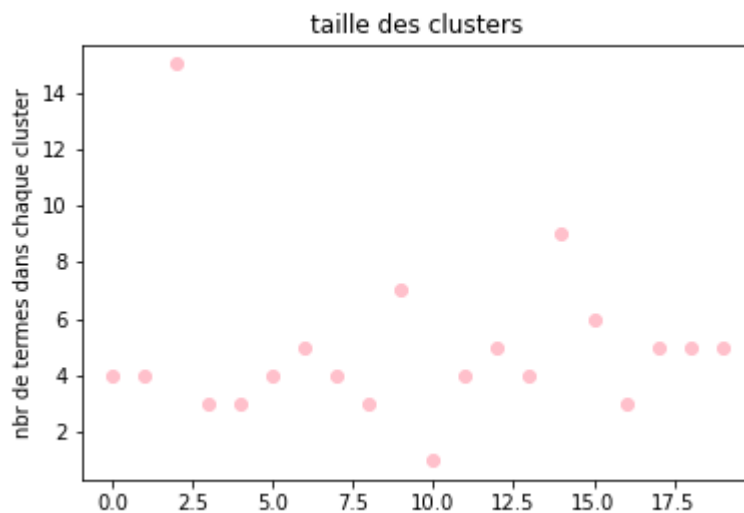
TD 3, rapport écrit
Amina Ahamada

- **Contenu du clusters**

```
},  
"ID 19": {  
  "Centro\u00efde": "JEANNE",  
  "Freq. centroide": 53,  
  "Termes": [  
    "JEANNE",  
    "JEANNS",  
    "JEANNp",  
    "JKANNE",  
    "Jean",  
    "Jeanne",  
    "Jeanne\u201d",  
    "Jeanno",  
    "Jeanre",  
    "Jeax",  
    "Jennne",  
    "Joanne",  
    "Saint-Jean",  
    "Vatan"  
  ]  
}
```

Ci-dessus, nous pouvons constater que les clusters ont été faits en fonction des suites de caractère en commun. Cela s'explique par l'usage de la distance cosinus.

- **Représentation graphique de la longueur des clusters**



Le nombre de termes dans les clusters varie entre 1 et 8, d'après le graphique le cluster le plus long contient 16 termes. Cependant, après vérification visuelle, j'ai constaté que le cluster le plus long a 13 termes.