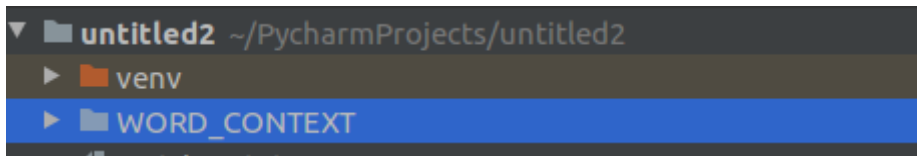


Suite au téléchargement de l'archive, veuillez ouvrir un nouveau projet sur l'ide pycharm et ajoutez-y le dossier WORD-CONTEXT (voir photo)



Ce dossier contient seulement les programmes, veuillez télécharger tous les modules nécessaires sur Pycharm en appuyant sur install requierement, un message qui s'affichera automatiquement sur votre ide ou bien en installant module par module.

Vous pouvez maintenant lancer les deux programmes principaux en allant sur le terminal de l'ide, situé en bas à gauche :

- Le premier avec la commande : `python WORD_CONTEXT/clustering/newk-means.py -p 'chemin du dossier contenant les mails' -f 'chemin du fichier .txt a comparé avec les mails'`
- Le deuxième avec la commande : `python WORD_CONTEXT/word2vec/interface.py`

Newk-means.py permet de comparer des mails afin de voir leur taux de similarité. Ce programme crée plusieurs fichiers en sortie : des fichiers matrices représentant les vecteurs de chaque mails et un fichier resultatcomparaison contenant le résultat de la comparaison entre les mails. (voir photo)

```
/home/etudiant/Bureau/pre/9_pre_processed.txt
/home/etudiant/Bureau/pre/4_pre_processed.txt
/home/etudiant/Bureau/pre/7_pre_processed.txt
/home/etudiant/Bureau/pre/6_pre_processed.txt
{'/home/etudiant/Bureau/pre/6_pre_processed.txt': '95.0%', '/home/etudiant/Bureau/pre/7_pre_processed.txt': '17.0%', '/home/etudiant/Bureau/pre/5_pre_processed.txt': '100.0%', '3.0%'}
```

Interface.py est une interface qui permet d'exécuter plusieurs programmes : (voir photo)

- Preprocess qui prend en entrée un fichier txt et permet de le traiter afin d'avoir un contenu propre pour les algorithmes.
- Word2vec permet d'exécuter l'algorithme word2vec sur un fichier txt
- Les 3 autres algorithmes k-means, hcluster et elbow prennent en entrée un fichier .bin, crée suite à l'exécution de word2vec

