

Série de TP 3 : Représentation du texte

Exercice 01 :

Word2Vec propose deux architectures :

- **CBOW (Continuous Bag of Words)** : prédit un mot cible à partir de son contexte.
- **Skip-gram** : prédit le contexte à partir d'un mot cible.

L'objectif de ce TP est de coder un modèle de Word2Vec utilisant la méthode CBOW avec **PyTorch** .

1. Implémentez la méthode preprocess() pour traiter le texte et générer le vocabulaire.
2. Implémentez generate_training_data() pour extraire les paires de contexte et de mot cible.
3. Implémentez la propagation avant (forward_pass) et la rétropropagation (backpropagation).
4. Entrez un texte d'entraînement
5. Testez le modèle CBOW en recherchant des mots similaires.
6. Expérimentez avec différents paramètres : window_size, embedding_dim, epochs, min_count.