



Similarité sémantique

Mesurer la similarité sémantique consiste à évaluer à quel point le sens des mots/phrases est proche

- > **Similarité entre mots**
- > **Similarité entre phrases**



Plusieurs approches pour calculer la similarité sémantique

-> Similarité vectorielle : deux mots sont similaires si leurs vecteurs sont similaires (vecteur tf-idf, embedding, ..etc)



Similarité vectorielle entre mots : méthode

- Retenir un modèle de représentation (word2vec, ..) -> Partie présentée par Célya
- représenter chaque mot sous forme d'un vecteur
- calculer la similarité vectorielle entre vecteurs de mots

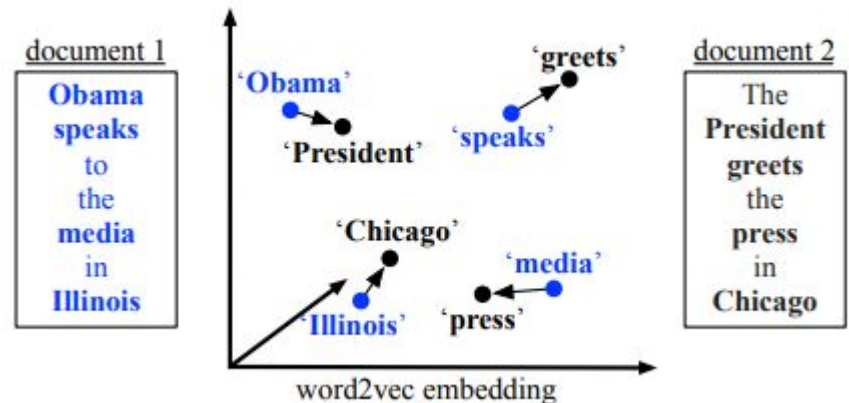
Similarité vectorielle entre phrases

- Choisir une méthode de calcul de similarité entre séquences de mots.

Calcul de la similarité

Plusieurs mesures de similarité :

- Distance euclidienne
- Produit scalaire
- Similarité cosinus
- Word Mover's Distance ----->





Utilité

- Détection du plagiat, la paraphrase
- Classer des documents (selon le sens)
- **Répondre à des requêtes : en reliant la requête à la réponse (texte -> résumé, titre ...) en considérant le sens**
- ...

Additionnel

- TP de M2 de calcul de similarités en utilisant différentes méthodes