

Lab Word Embedding n°2 part 1

Modélisation

Modélisation d'un mot

1. Modélisant un mot par un vecteur d'entiers (ou chaque entier représente une lettre unique)

exemple : alpha $\rightarrow (0,11,15,7,0)$

Soit E l'ensemble de tous les mots.

Définissons la classe **Mot**.

2. Soit l'opération concaténation définie de la façon suivante :

$$\begin{aligned} \cdot & : E \times E \rightarrow E \\ (a_1, \dots, a_n), (b_1, \dots, b_n) & \rightarrow (a_1, \dots, a_n, b_1, \dots, b_n) \end{aligned}$$

3. A partir de l'opération concaténer, écrire les fonctions `accorder_au_feminin` (qui consiste à ajouter un e à la fin d'un mot au masculin) et `accorder_au_pluriel` (qui consiste à ajouter un s à la fin d'un mot au singulier).
4. Réécrire la fonction `repr` de la classe **Mot** qui retourne la représentation textuelle du mot et la classe `toVect` qui retourne sa représentation vectorielle.

Modélisation d'une expression

5. Dans cette partie nous allons modéliser les expressions.
Commençons par les noms communs.

Modélisons les noms communs par un couple (d,n) où d est la quantité (1,2,...) et n le mot.

exemple : (1, pomme), (2, cours)

Définissons la classe **Expression** de laquelle hérite la classe **NomCommun**.

6. Réécrire la fonction repr de **NomCommun** qui, à partir du couple (d,n), retourne la représentation textuelle accordée selon le genre et le nombre.

exemple : (3, pomme) → Trois pommes

Modélisation d'une phrase

7. Afin de simplifier la modélisation des phrases, nous allons modéliser les phrases les plus courtes.

Sujet + Verbe : Le sujet étant un nom commun est le verbe étant un mot (le verbe sera représenté par sa forme infinitive).

Définir donc la classe Phrase.

8. Réécrire la méthode repr de la classe **Phrase** qui retourne la représentation textuelle de cette phrase accordée.

exemple : ((3, pomme), tomber) → Trois pommes tombent

Afin d'achever cette opération, un retour vers la classe **Mot** s'impose pour définir la fonction de conjugaison (On se limitera aux verbes du premier groupe)

Base de données et 1er test

9. Maintenant que les models ont été posés et que quelques opérations ont été définies, il s'agit de remplir notre base de données de noms et de verbes. Téléchargez le fichier json suivant :
<https://github.com/AttilaDSA/IntilagDSAcademy/blob/master/Word%20Embedding%20Labs/Word%20Embedding%20Lab%20n%C2%B02/dictionary.rar>
10. Extraire du fichier json les noms et les verbes du 1^{er} groupe.
11. Ecrire un petit code qui choisit un chiffre, un mot et un verbe au hasard et qui en construit une phrase bien accordée.