

Exercice IA GL Code2Vec

https://github.com/lailachm/IA_GL

J'ai travaillé sous windows, sur l'IDE PyCharm mais je n'ai pas pu compléter cet exercice. En effet, à partir de la step 4, une erreur m'empêche de continuer.

```
PS C:\Users\Laïla\code2vec> python3 code2vec.py --load models/java14_model/saved_model_iter8.release --predict
Traceback (most recent call last):
  File "C:\Users\Laïla\code2vec\code2vec.py", line 1, in <module>
    from vocabularies import VocabType
  File "C:\Users\Laïla\code2vec\vocabularies.py", line 7, in <module>
    import tensorflow as tf
ModuleNotFoundError: No module named 'tensorflow'
PS C:\Users\Laïla\code2vec>
```

En tapant la commande de la step4, j'ai une erreur m'indiquant que le module tensorflow n'est pas trouvé.

Or, avec la commande `pip show tensorflow`, j'obtiens le résultat suivant :

```
PS C:\Users\Laïla\code2vec> pip show tensorflow
Name: tensorflow
Version: 2.5.0
Summary: TensorFlow is an open source machine learning framework for everyone.
Home-page: https://www.tensorflow.org/
Author: Google Inc.
Author-email: packages@tensorflow.org
License: Apache 2.0
Location: c:\users\laïla\code2vec\venv\lib\site-packages
Requires: wrapt, wheel, protobuf, astunparse, keras-nightly, gast, keras-preprocessing, grpcio, numpy, opt-einsum, tensorboard, termcolor, h5py, tensorflow-estimator, six, flatbuffers, absl-py, typing-extensions, google-pasta
Required-by:
PS C:\Users\Laïla\code2vec>
```

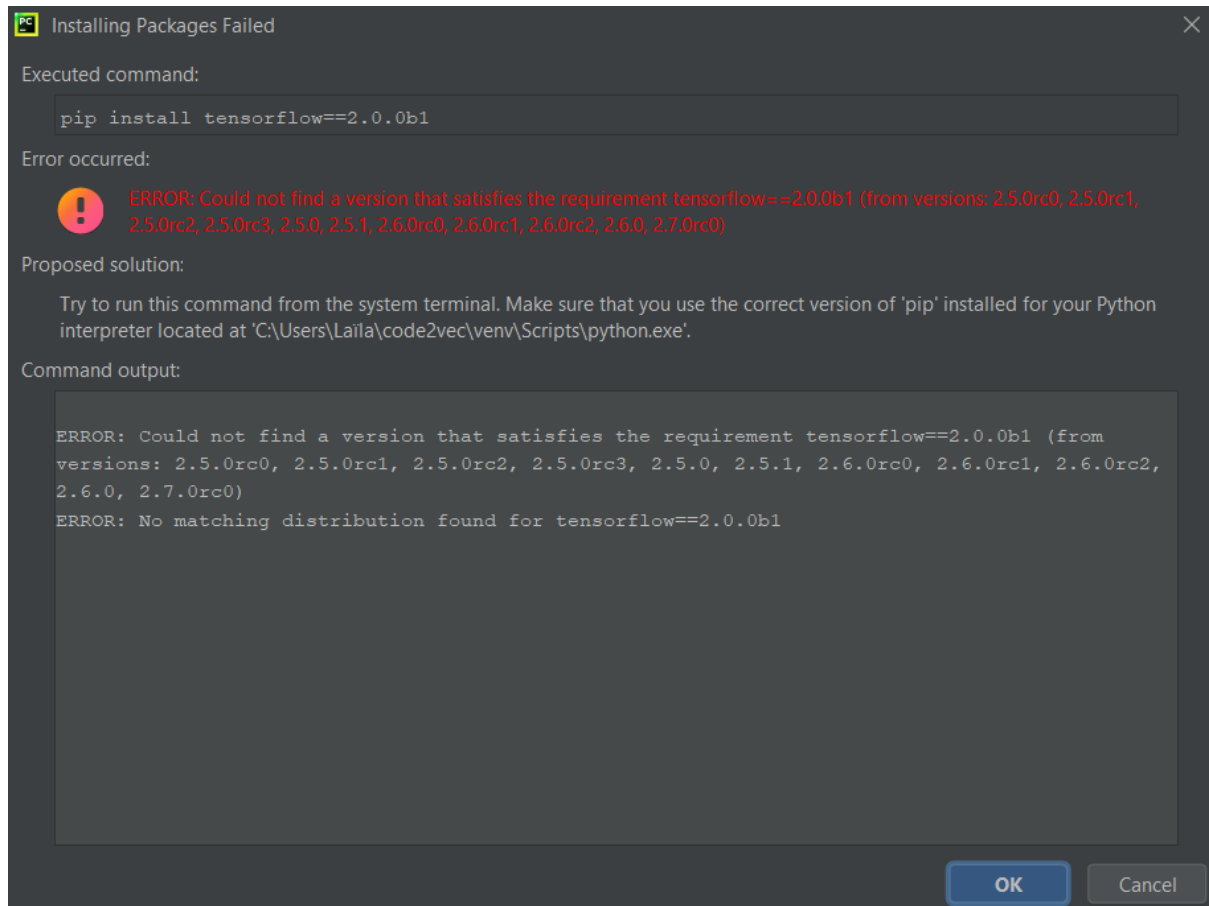
Cela indique que tensorflow est bien installé (version 2.5.0)

Je l'ai également dans mes packages

users>Laïla>code2vec>venv>Lib>site-packages :

PC > OS (C:) > Utilisateurs > Laïla > code2vec > venv > Lib > site-packages					▼	🔄
Nom		Modifié le	Type	Taille		
📁	six-1.15.0.dist-info	10/10/2021 16:55	Dossier de fichiers			
📁	tensorboard	10/10/2021 16:55	Dossier de fichiers			
📁	tensorboard_data_server	10/10/2021 16:55	Dossier de fichiers			
📁	tensorboard_data_server-0.6.1.dist-info	10/10/2021 16:55	Dossier de fichiers			
📁	tensorboard_plugin_wit	10/10/2021 16:55	Dossier de fichiers			
📁	tensorboard_plugin_wit-1.8.0.dist-info	10/10/2021 16:55	Dossier de fichiers			
📁	tensorboard-2.6.0.dist-info	10/10/2021 16:55	Dossier de fichiers			
📁	tensorflow	10/10/2021 18:21	Dossier de fichiers			
📁	tensorflow_estimator	10/10/2021 18:21	Dossier de fichiers			
📁	tensorflow_estimator-2.5.0.dist-info	10/10/2021 18:21	Dossier de fichiers			
📁	tensorflow_gpu-2.5.0.dist-info	10/10/2021 18:14	Dossier de fichiers			
📁	tensorflow-2.5.0.dist-info	10/10/2021 18:21	Dossier de fichiers			
📁	termcolor-1.1.0-py3.9.egg-info	10/10/2021 16:55	Dossier de fichiers			

Je note également que dans le fichier `code2vec>requirements.txt`, la version requise pour le module tensorflow est 2.0.0-beta1, mais cette version n'est pas téléchargeable. Voici l'erreur qui s'affiche :



Voilà pourquoi la version de tensorflow qui est montrée plus haut est une version ultérieure (2.5.0).

Après plusieurs heures de recherches infructueuses, j'ai fini par me résoudre à répondre aux questions en utilisant le site <https://code2vec.org/> afin de visualiser le résultat du projet, en espérant que cela convienne. Merci de votre compréhension.

Trouver un moyen de récupérer de façon programmatique les context-paths et leurs poids (attention vector).

Pour récupérer les context-paths et leur poids qui s'affichent après la Step 4, il faut soit récupérer le contenu affiché en terminal directement, soit récupérer les données sous forme de structure Python. Pour cela, une fois le projet code2vec lancé, le contenu du terminal peut être transféré dans un fichier, via une commande Shell (sous windows) par exemple :

```
python3 code2vec.py --load  
models/java14_model/saved_model_iter8.release --predict | Out-File  
-FilePath C:\Users\Laïla\result.txt
```

Est-ce pertinent d'utiliser ces paths pour identifier des bad-smells si par exemple il n'y a pas de poids dominant (méthode avec plusieurs fonctionnalités)?

Dans le cas où il n'y a pas de poids dominant, ces paths seraient utiles pour repérer du code qui se répète, et qui pourrait donc être factorisé. Si on prend l'exemple d'un projet qui comporte plusieurs classes qui implémentent les mêmes méthodes, ces paths pourraient mettre en avant le code dupliqué, et indiquer qu'une interface serait utile.