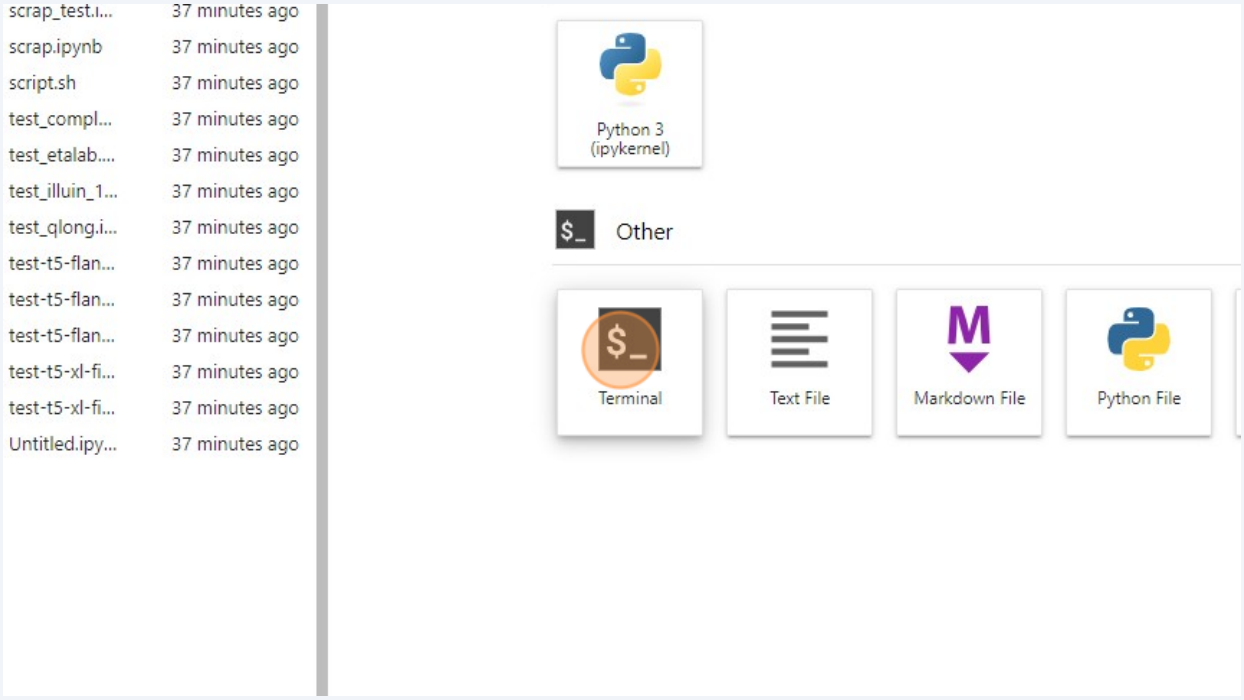
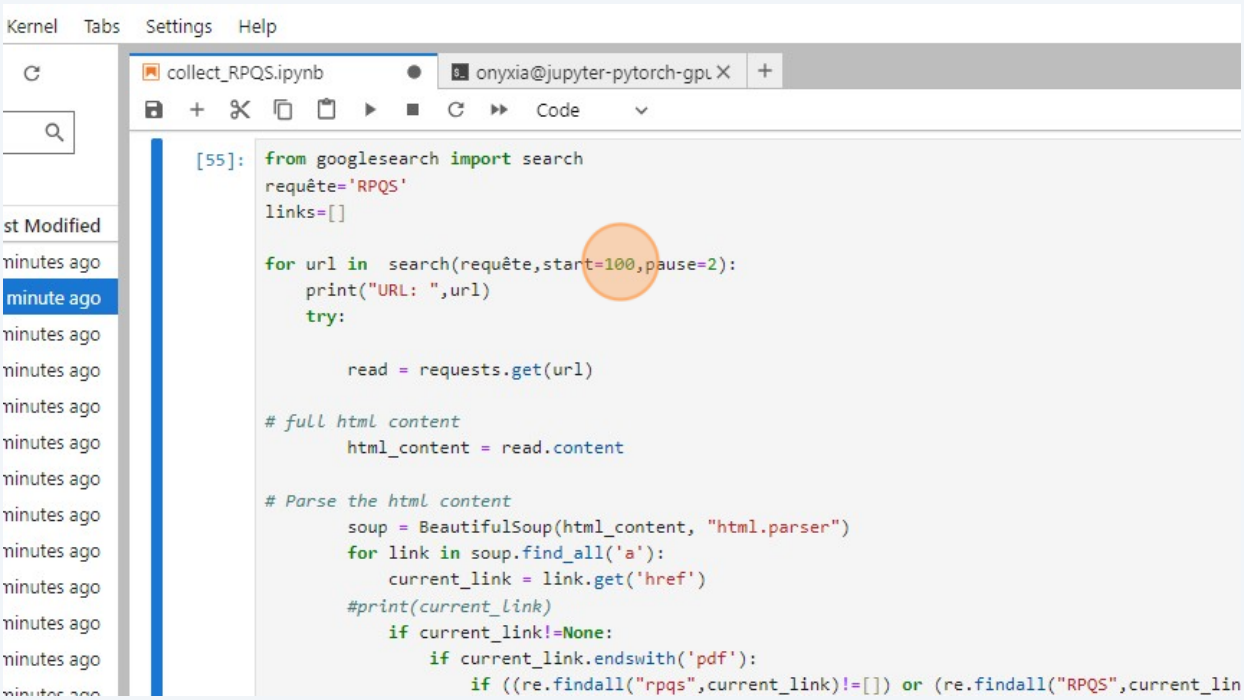
Exécuter le code de collecte de PDF en ligne

1 ouvrir une fenêtre de terminal dans une instance de jupyter-pytorch-gpu sur le ssp cloud.

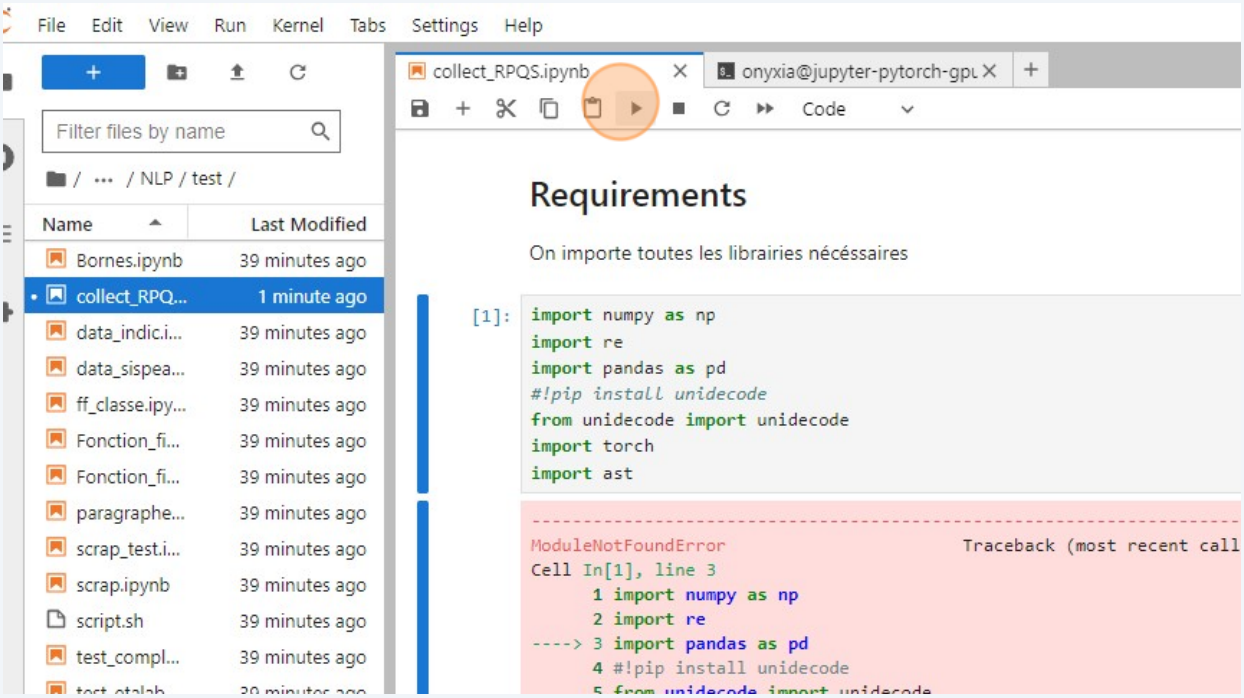


2 Type "cd/sispea/NLP" puis "pip install -r requirements.txt" pour intsaller les librairies nécessaires

3 Modifier le numéro "start", qui représente le premier lien à récupérer en ligne. La commande search va par défaut retourner 100 liens.



4 Exécuter toutes les cellules du notebooks



5 Le fichier csv final avec les liens des PDF sera enregistré sur le dépôt s3