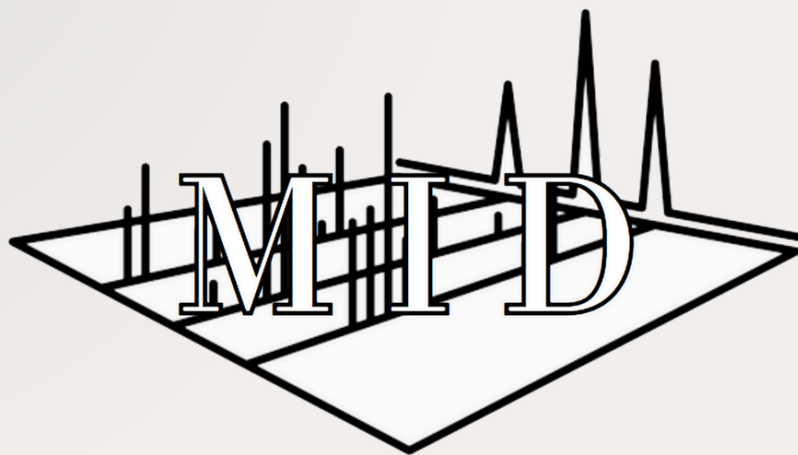


Molecule_ID based MSMS
fingerprint Similarity



AJOUTER UN
SPECTRE A LA
BASE DE
REFERENCE

Fayçal Hassani

SOMMAIRE

01. EXPLIQUATION

02. AJOUT DU SPECTRE

Explication

Pour obtenir un modèle aussi fiable que possible, il est crucial de disposer d'une large base de référence.

Pour ce faire, il est nécessaire de convertir un fichier RAW en format mzML.

Une fois cette conversion effectuée, il faut ensuite ajouter ce spectre au dictionnaire.

Cette étape permet d'enrichir la base de données et d'améliorer la précision et la fiabilité du modèle.

ajout du spectre

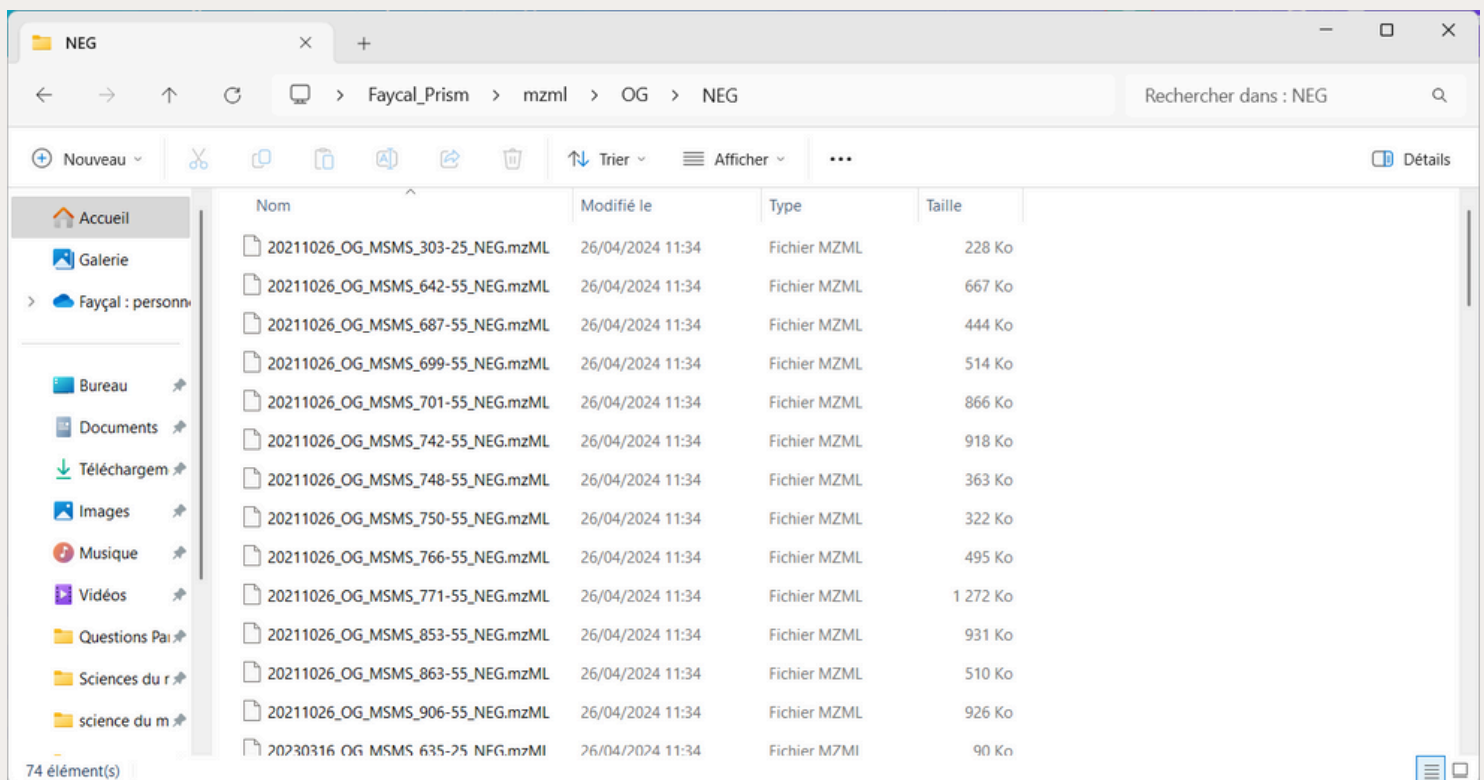
Pour convertir des fichiers de données brutes en format mzML, il est nécessaire d'ouvrir l'invite de commande et d'utiliser la commande suivante : **msconvert "\$raw_file" --mzML -o "\$output_directory"**.

Cependant, avant de pouvoir exécuter cette commande, il est impératif d'installer ProteoWizard.

Pour ce faire, il est recommandé de consulter le manuel d'utilisation afin de s'assurer que l'installation est correctement effectuée et que toutes les dépendances nécessaires sont satisfaites.

La prochaine étape consiste à déplacer ce fichier nouvellement créé vers le répertoire où sont stockés tous les autres fichiers mzML.

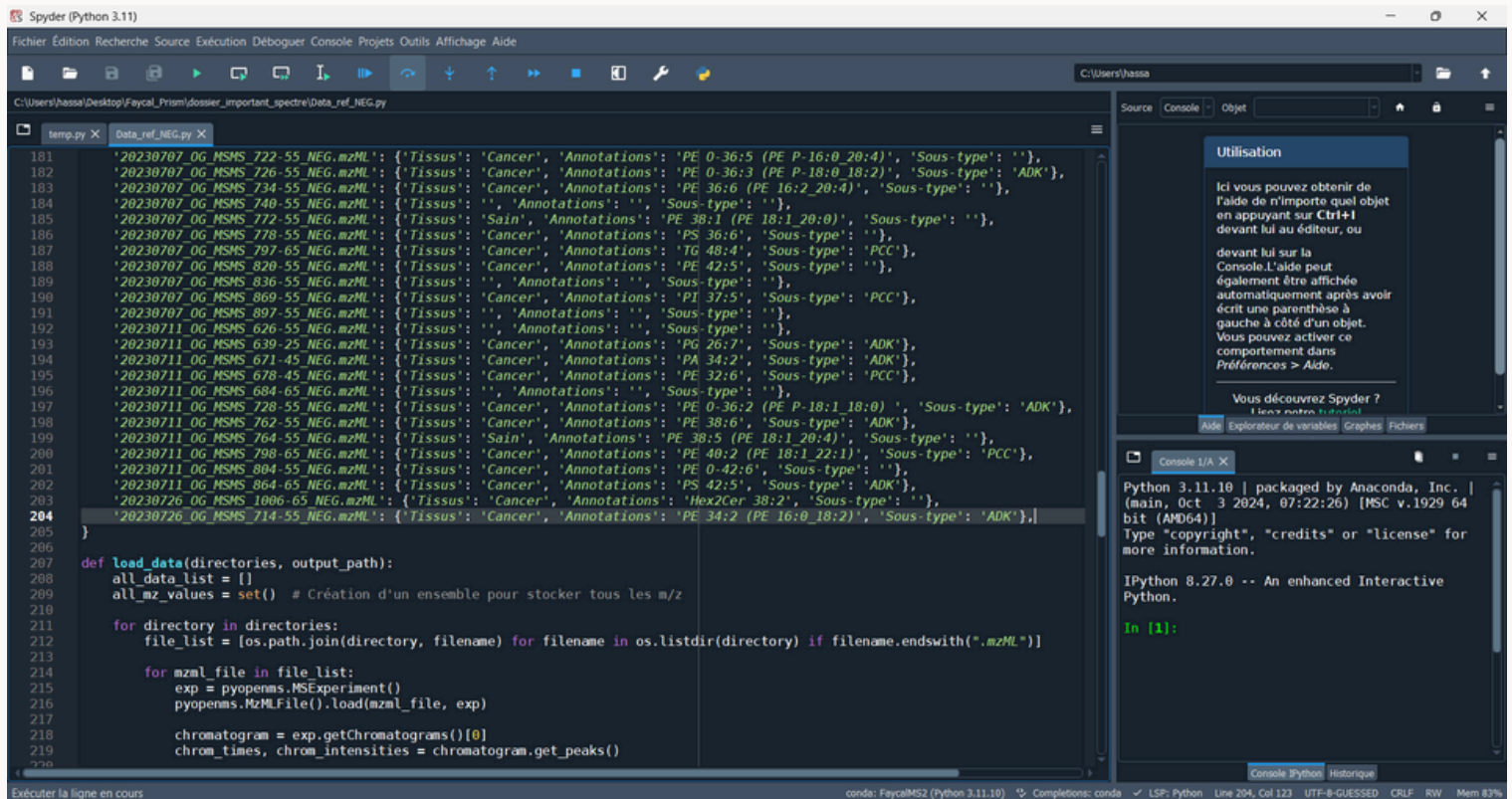
Cela peut être fait en glissant simplement le fichier converti dans le dossier approprié.



Cette organisation permet de centraliser tous les fichiers mzML dans un seul emplacement, facilitant ainsi leur gestion et leur analyse ultérieure.

Maintenant, il faut venir ajouter une phrase au dictionnaire en suivant ce format :

**"nomdufichier.mzML": {'Tissus': '',
'Annotations': '', 'Sous-type': ''}.**



```
181 '20230707_06_MSMS_722-55_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PE 0-36:5 (PE P-16:0_20:4)', 'Sous-type': ''},
182 '20230707_06_MSMS_726-55_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PE 0-36:3 (PE P-18:0_18:2)', 'Sous-type': 'ADK'},
183 '20230707_06_MSMS_734-55_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PE 36:6 (PE 16:2_20:4)', 'Sous-type': ''},
184 '20230707_06_MSMS_740-55_NEG.mzML': {'Tissus': '', 'Annotations': '', 'Sous-type': ''},
185 '20230707_06_MSMS_772-55_NEG.mzML': {'Tissus': 'Sain', 'Annotations': 'PE 38:1 (PE 18:1_20:0)', 'Sous-type': ''},
186 '20230707_06_MSMS_778-55_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PS 36:6', 'Sous-type': ''},
187 '20230707_06_MSMS_797-65_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'TG 48:4', 'Sous-type': 'PCC'},
188 '20230707_06_MSMS_820-55_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PE 42:5', 'Sous-type': ''},
189 '20230707_06_MSMS_836-55_NEG.mzML': {'Tissus': '', 'Annotations': '', 'Sous-type': ''},
190 '20230707_06_MSMS_869-55_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PI 37:5', 'Sous-type': 'PCC'},
191 '20230707_06_MSMS_897-55_NEG.mzML': {'Tissus': '', 'Annotations': '', 'Sous-type': ''},
192 '20230711_06_MSMS_626-55_NEG.mzML': {'Tissus': '', 'Annotations': '', 'Sous-type': ''},
193 '20230711_06_MSMS_639-25_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PG 26:7', 'Sous-type': 'ADK'},
194 '20230711_06_MSMS_671-45_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PA 34:2', 'Sous-type': 'ADK'},
195 '20230711_06_MSMS_678-45_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PE 32:6', 'Sous-type': 'PCC'},
196 '20230711_06_MSMS_684-65_NEG.mzML': {'Tissus': '', 'Annotations': '', 'Sous-type': ''},
197 '20230711_06_MSMS_728-55_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PE 0-36:2 (PE P-18:1_18:0)', 'Sous-type': 'ADK'},
198 '20230711_06_MSMS_762-55_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PE 38:6', 'Sous-type': 'ADK'},
199 '20230711_06_MSMS_764-55_NEG.mzML': {'Tissus': 'Sain', 'Annotations': 'PE 38:5 (PE 18:1_20:4)', 'Sous-type': ''},
200 '20230711_06_MSMS_798-65_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PE 40:2 (PE 18:1_22:1)', 'Sous-type': 'PCC'},
201 '20230711_06_MSMS_804-55_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PE 0-42:6', 'Sous-type': ''},
202 '20230711_06_MSMS_864-65_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PS 42:5', 'Sous-type': 'ADK'},
203 '20230726_06_MSMS_1006-65_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'Hex2Cer 38:2', 'Sous-type': ''},
204 '20230726_06_MSMS_714-55_NEG.mzML': {'Tissus': 'Cancer', 'Annotations': 'PE 34:2 (PE 16:0_18:2)', 'Sous-type': 'ADK'},
205 }
206
207 def load_data(directories, output_path):
208     all_data_list = []
209     all_mz_values = set() # Création d'un ensemble pour stocker tous les m/z
210
211     for directory in directories:
212         file_list = [os.path.join(directory, filename) for filename in os.listdir(directory) if filename.endswith(".mzML")]
213
214         for mzml_file in file_list:
215             exp = pyopenms.MSExperiment()
216             pyopenms.MzMLFile().load(mzml_file, exp)
217
218             chromatogram = exp.getChromatograms()[0]
219             chrom_times, chrom_intensities = chromatogram.get_peaks()
220
221
```

Ici, vous pourrez préciser et donner des informations supplémentaires telles que le type de tissu, les annotations pertinentes, et le sous-type spécifique du fichier mzML.

Il suffit maintenant de lancer le programme pour obtenir le fichier Parquet, qui servira de base de référence pour toutes les analyses ultérieures.