**Overview of cellParser.py**

The function of the script in cellparser.py is to generate a file [cells\_data.csv] which can be used by batchneuroml.py to insert the details of cells into database.

The required files for generation of cells\_data.csv are:

1. All the cell files which are with extension of. **nml.**
2. The file term\_mapping.csv , which is required to map the Term in cells to the keywords, neurolexTerm and NeurolexURI.

The generated file (cells\_data.csv) have the columns [modelType, modelName, fileName, children, references, neurolexTerm, neurolexURI, keywords, pubmedID, translator, authors].

modelType :- As we are inserting cells modelType is CL.

modelName : Name of the file without cell.nml

filename : Name of the file with extension .cell.nml

children references : All the channels that the cell is referencing

references : - The project from where the cells data is taken from.

neurolexTerm, neurolexURI, keywords :- Taken from file **term\_mapping.csv**  written by Professor Sharon.

|  |  |
| --- | --- |
| **Function** | **Description** |
| \_\_main\_\_ | program initialization:   * Initiates the process. * Processes the data for each cell. * Writes to the cells\_data.csv * Checks if any cells is not related to channel. |
| Readfiles() | Checks the folder path where all cells are present. Takes all the files with .nml extension and adds the filename to content\_list array. |
| getCellNamesFromLems(): | * Iterates through all the files in content\_list array.  Gets all the channels with which the cell is associated and adds it to child cells array. * Extracts the tern from the filename which is after the level [DPC,BAC, BP etc.] * Calls the function **getneurolexId** . to get the values of keywords, neurloexTerm and neurolexURI. * Adds the data to children set variable. * Sends the entire data to getChannelsFromFiles which adds the data for the cell to excel sheet [cells\_data.csv] |
| addToFile | * Takes input of all the details from getChannelsFromFiles function and writes it to the **cells\_data.csv** |
| getChannelsFromFiles | * Takes the input of all the data related to a cell from **getChannelsFromFiles** . * Puts together the received data into row. * Calls the addToFile to write the row into excel file [**cells\_data.csv**]. |
| getThreeDigitTerm | * Takes the filename as input * Invoked from **getCellNamesFromLems**  function * Splits the filename based on (“\_”) and returns the Term |
| getneurolexId | * Takes the 3 digit tern as input * Invoked from **getCellNamesFromLems.** * Reads the content from **term\_mapping.csv** * Returns null if the term is not present || Returns the data if term is present |
| getChildrenFromChannels | * Reads the file **ChannelsReformat.csv** and returns all the channels present in the file. * This is to check if there are any extra cells without channels * Adds all the data to channelChildren array. |
| nonExistingChannels | * Returns if any cells is pointing to non Existing Channel. * **Please check with professor if Data is Present Here** |
| getPubMedId | * Returns the PubMedId. * Invoked from **getChannelsFromFiles** function. * As all the cells are taken from single research paper. We have directly given the Pubmed number. * Please re-write the function as needed. |