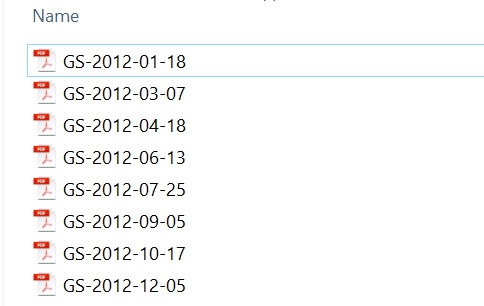
# **NLP – PDF Text Search (Word Occurances)**

**Step 1** – Store all PDF files in accessible windows folder



**Text Preprocessing** – Removed Punctuations, [Stopwords](https://pythonspot.com/nltk-stop-words/)

**Step 2** – Run the python file from command line with two parameters

Parameter 1 – Folder Path

Eg C:\AI\_Demo\NLP\_Demo\GreenSheets-2012

Parameter 2 – Word or phrases to be searched

Searchword = Percent Or ‘Percent Change’

Command line Argument :



Output :

List of Filenames and Count of “search word” in that file

