System: OS X, Windows 10
Language: Python 2.7.6
External Libraries: json, nltk
Tool: Sublime text2, Pycharm 5.0.1

Files (folders):
1) 1-readme.txt - contains instructions to compile and run programs.

2) 2-source_code - contains Python program for all tasks and needed files.

Phase 1:

Task1(According to Professor's answer in piazza, we didn't attach lucene.java):
* BM25.py - BM25 search engine
* tf_idf.py - tf_idf search engine

Task2:
* Thesauri.py - query expansion using thesauri
* Dice.py - query expansion using Dice's formula

Task3:
* BM25_Stop.py - BM25 search engine removing stop words
* bm25_Stem.py - BM25 search engine for stemmed queries/corpus

Phase 2:

* tf_idf_Stop.py - tf_idf search engine removing stop words
* eval_write_xls.py - the program to generate excel table of evaluation for each search engine.

Extra Credits:

* Snippet.py - generate snippet for each query result and highlight query terms

NOTE: Please CHANGE the file name/path in codes for each search engine when you download them into your computer if necessary.

3) 3-report.txt - a short report describing my implementation.

4) Folders named BM25_SE, Lucene_SE, Tf_idf_SE contains top 100 ranks for each search engine for task 1 in Phase 1.

5) Folders named BM25_QE_Thesauri_SE, BM25_QE_Dices _SE contains top 100 ranks for BM25 search engine using query expansion for task 2 in Phase 1.

6) Folders named BM25_Stop_SE, BM25_Stem contains text file(s) of top 100 ranks for BM25 search engine using common_words(stop words list) or stemmed queries/documents for task 3 of Phase 1.

7) Tf_idf_Stop_SE contains top 100 ranks for tf_idf search engine using common_words(stop words list) for Phase 2(2).

8)
- Folders named BM25_eval, Tf-idf_eval, Lucene_eval contains evaluation tables for search engines in task 1 of Phase 1.
- Thesauri_eval, Dice_eval contains evaluation tables for search engines in task 2 of Phase 1.
- BM25_Stop_eval contains evaluation tables for search engine in task 3(A) of Phase 1.
- Tf_idf_Stop_eval contains evaluation tables for search engine in task 1 of Phase 2.

9) In folder named Evaluation, SE_AVP_RR.xlsx contains AP and PP values for each queries in different search engine. SE_Eval contains MAP and MRR for each search engine.

10) For extra credits part, BM25_SE_Snippet, Tf_idf_SE_Snippet, Lucene_SE_Snippet, BM25_Stop_SE_Snippet, BM25_QE_Thesauri_SE_snippet, BM25_QE_Dices_SE_snippet, Tf_idf_Stop_SE_Snippet are presented for 7 different runs.

How to install libraries:
1) For nltk, download setuptools in "https://pypi.python.org/simple/setuptools/", run "sudo sh Downloads/setuptools-0.6c11-py2.7.egg", "sudo pip install --ignore-installed six -U numpy", and "sudo pip install --ignore-installed six -U pyyaml nltk" one by one in terminal.
Note: If this is your first time to use nltk, run "import nltk" AND "nltk.download('punkt')" in you IDE to download necessary data package.

2) Json is generally included in versions after python 2.6, if not, please download it, too.

Two ways to run the program:
1) Cd into the directory ~/FolderName/, and run "python filename.py".
2) Use sublime to run these programs, press cmd+b.
3) Before running the program, please change the file path or name in the code to read/output data to specific folders