Abstract Text Analysis Process:

Step 0 : set up the Python environment by installing python 2.7, pip and all python libraries outlined in the dependencies.txt file inside /src/data\_analysis/pythonCode. Do make this easier, you should run the pythonLib\_setUp.sh shell script found in /src/data\_analysis/pythonCode.

1. DocSim: Navigate to /src/data\_analysis/pythonCode/ and run the command:

**DocSim.py -i "../../../data/conference.csv" -c 2**

1. Doc2Vec\_pretrained 1: Navigate to /src/data\_analysis/pythonCode/ and run the command:

**Doc2Vec\_pretrained.py -i "../../../data/conference.csv" -c 2 –m "../languageModels/apnews\_doc2vec.bin"**

1. Doc2Vec\_pretrained 2: Navigate to /src/data\_analysis/pythonCode/ and run the command:

**Doc2Vec\_pretrained.py -i "../../../data/conference.csv" -c 2 –m "../languageModels/enwiki\_doc2vec.bin"**

1. Doc2Vec\_selfTrained: Navigate to /src/data\_analysis/pythonCode/ and run the command:

**Doc2Vec\_selfTrained.py -i "../../../data/conference.csv" -c 2**

1. WMD: Navigate to /src/data\_analysis/pythonCode/ and run the command:

**WMD.py -i "../../../data/conference.csv" -c 2 –m "../languageModels/WMD\_model\_withGoogleNews\_W2V.bin"**

1. InferSent (Only running on Linux or OSX, because of library limitations (i.e.: pytorch)):

Run main.py inside the InferSent folder from a Linux or OSX operating system, after pytorch was configured. Adjust file path inside main.py, if needed.

Abstract Clustering Process:

1. Set R working directory to the root directory of this project’s directory, i.e. SSC Consulting, so for example run the R command “setwd("C:/Users/Norbert/Desktop/SSC consulting")”.
2. Rerun whichever model’s clustering analysis. Careful with the script, some of it is just exploring the data, to decide on the proper number of clusters within the data.
3. In order to perform cross validation or stability checking, go to tests/data\_analysis/clustering\_R and run Clustering Analysis Validation using CV.R or Clustering Analysis Validation using clValid package.R