Project Progress Report for 2023 Fall CS410: Course Project
Topic: Implementation of Compression Text Classification Method
Team member: Kiyotaka Kokubun (kokubun3@illinois.edu)

In the progress report of November 20th, I have been keeping continuing planning on how I can perform the compress text classification. One of possible ways is that I will use BM25 and compress method for the Cranfield dataset and calculate the score. I may change it from metapy to more general text data sets available when I face any difficulties to calculate performance in the same environment.
I have implemented the compress classification function as a function in python from the original paper, and I will be writing a function which takes Cranfield dataset and calculate the score in the same way to metapy.

- Additional comment on December 1st
    - I did not use metapy at all since the scoring method is not fully opened. I instead wrote scoring code manually by referring the score and index from Cranfield dataset as well as nDCG calculation and executed BM25Okapi from rank_bm25 library.
    - After implementing compressed method for text retrieval function with Cranfield dataset and compered to BM25, it is now very difficult to use this compression method for text retrieval. The NDCG scores are generally 10 times lower than ones by BM25.
    - Therefore, I decided not to create search engine app using this method, and I will work on a couple of notebook to introduce its official brand-new package for nearest neighbors text classification.