Problem description:

1. Find 100 CNN news articles online (try to find them in different categories, e.g., sports and finance). You need to find the new article by yourself. Pls ignore picture or other non-text data in the new article.
2. Convert them to data matrix (each row is an article and each column is a unique term).
3. Run K-means clustering with Euclidean, Cosine and Jarcard similarity. (Specify K as the number of categories of your 100 CNN news articles)
4. Evaluate K-means clustering results with SSE
5. Identify and interpret your findings about experimental results

Deadline: please submit your results by 02/27 . You may compress your 100 news articles, data matrix, K-means results and your own discussion into a zip file to submit.