HW4

Here we offer the dataset, Google AI published Research, which is crawled from(https://ai.google/research/pubs/).In this dataset, we only offer 'title' and 'abstract' and concatenate both (Google_AI_published_research.csv).

1. Preprocessing of dataset (Transforming the text instances into a tokenized word vector matrix which is an matrix for demonstrating the contents in D documents with v word. Each row represents a document instance while each column stands for a selected word)

2. In this homework assignment, we will need to use five methods to cluster. Note that method∈{LDA, Agglomerative, KMeans, KMeans++, FCM}.

3. How do you select the parameters?

4. Note that the number of clusters must be greater than 2.