Bài toán phân cụm văn bản (với tập dữ liệu đã lấy trong tuần 2)
Sử dụng thuật toán phân cụm văn bản để xác định xem tập văn bản đầu vào đó chia được thành mấy cụm. Với tập đầu vào bạn đã chủ động lấy mẫu sẵn theo 2 cụm thì áp dụng thuật toán phân cụm k-mean với k=2 trước
Sau đó test lại thuật toán với số cụm k=3,5 và đánh giá độ ổn định của các cụm (tính theo tổng khoảng cách trung bình của tất cả các phần tử trong cụm)

Khoảng cách giữa các phần tử bạn có thể dùng theo độ đo tương tự cosin áp dụng cho văn bản

Chú ý:

- Nên dùng thư viện để cài đặt thuật toán phân cụm văn bản