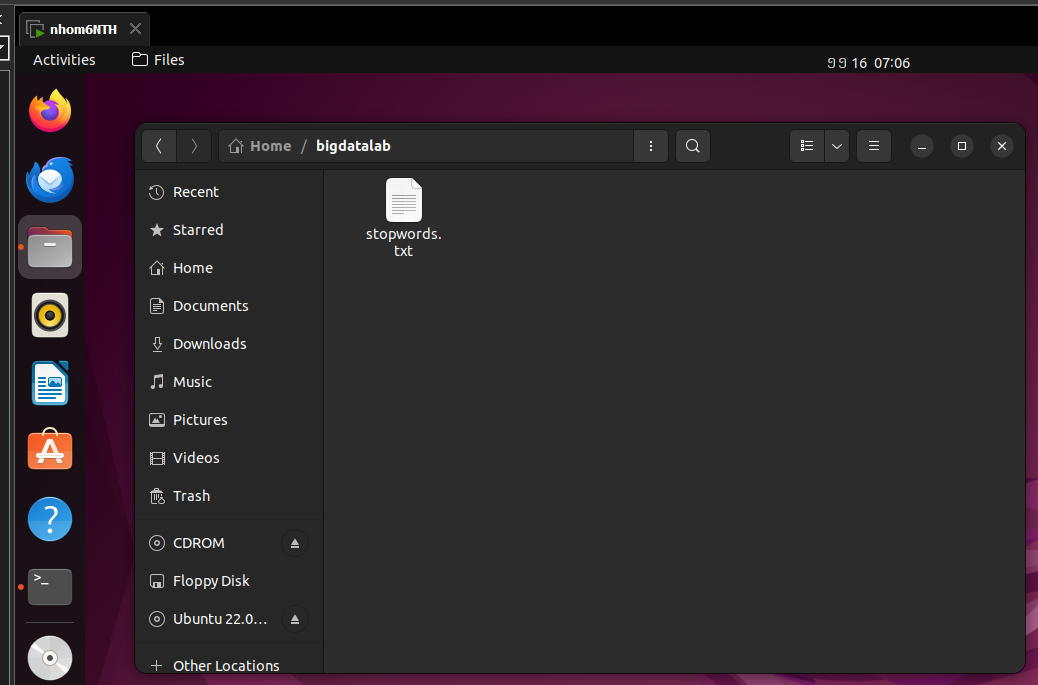
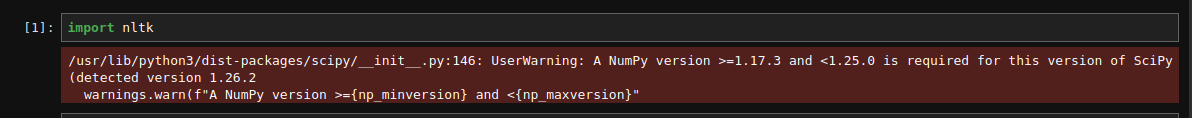
**Lab4**

B1: Tạo tệp stopwords.txt

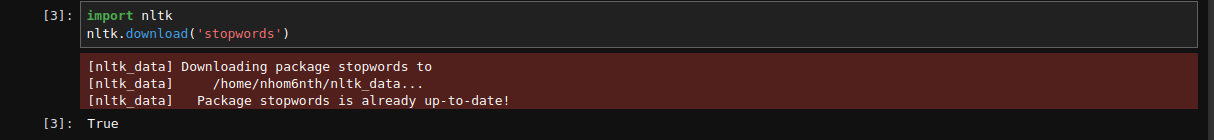


B2: khai báo thư viện nltk

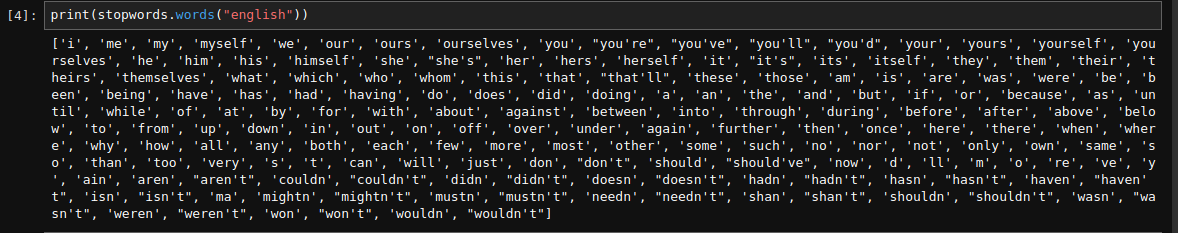


B3: import module stopwords từ thư viện NLTK

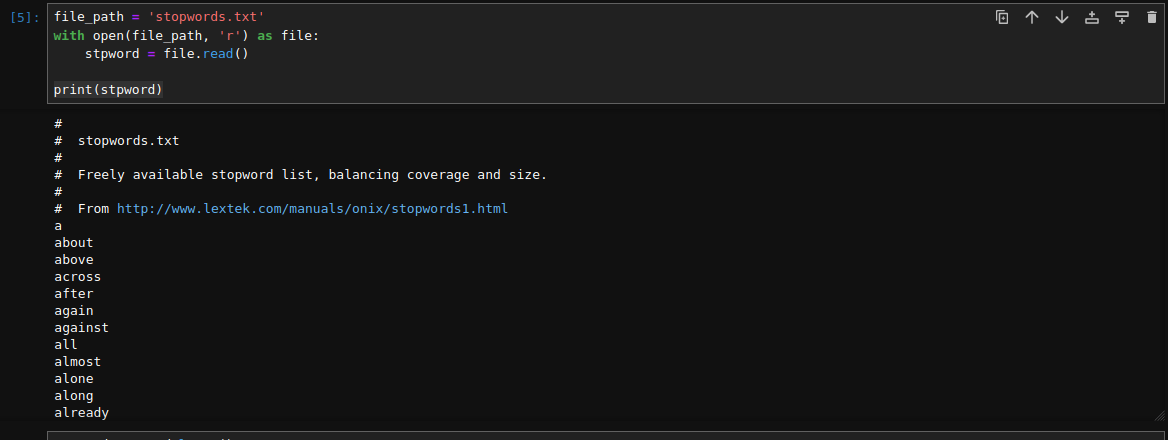


B4: tải tập tin stpwords từ thư việc nltk  


B5: in ra danh sách các từ ngừng (stopwords) trong tiếng anh



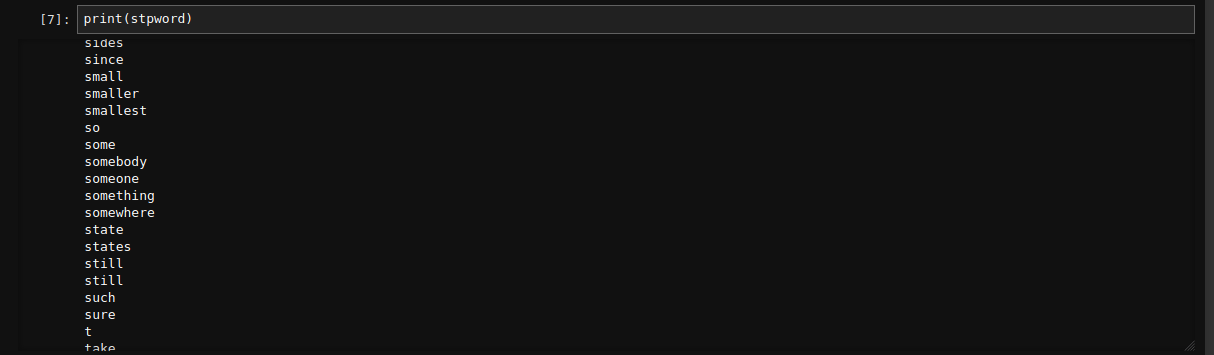
B6: đọc tệp stopwords.txt gán thành tên stpword



B7: chuyển đổi tất cả các ký tự trong stpword thành chữ thường (lowercase)



B8: in ra stpword vừa chuyển đổi

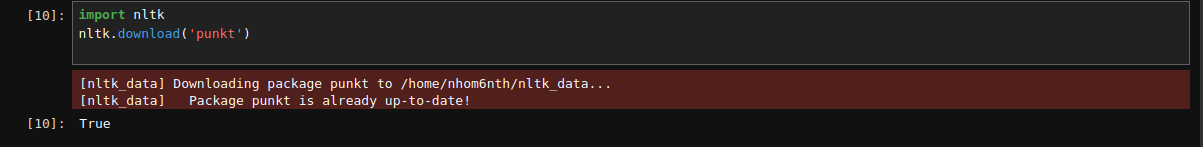


B9: sử dụng thư viện NLTK để tạo một tập hợp set chứa các từ dừng stopwords trong tiếng Anh



B10: sử dụng thư viện NLTK import hàm word\_tokenize từ module nltk.tokenize

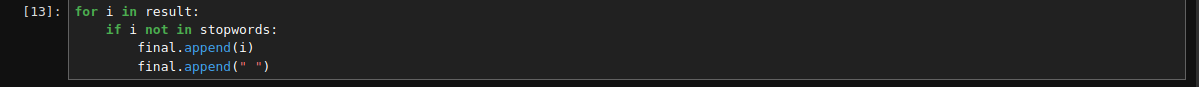
  
B11: tải punkt thư viện nltk



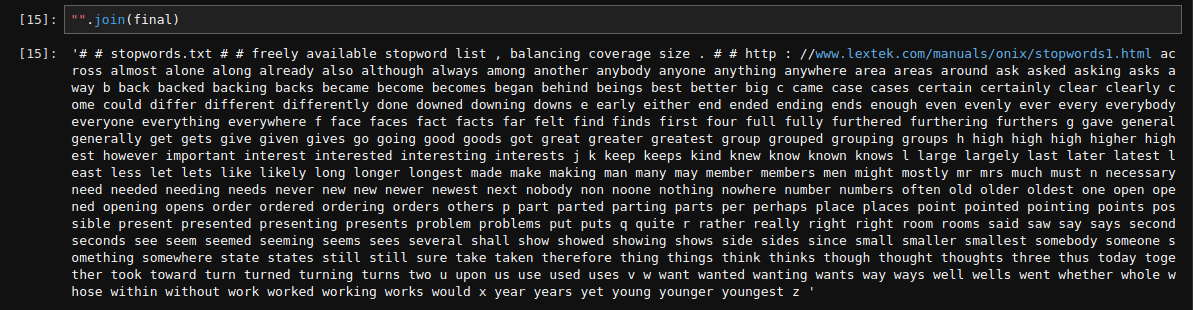
B12: sử dụng hàm word\_tokenize từ thư viện NLTK để phân tách từ trong một biến có tên là stpword



B13: Khởi tạo danh sách trống  


B14:   


B15: in ra kết quả

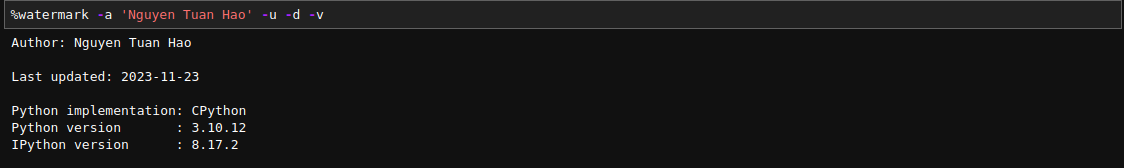


Bài 5

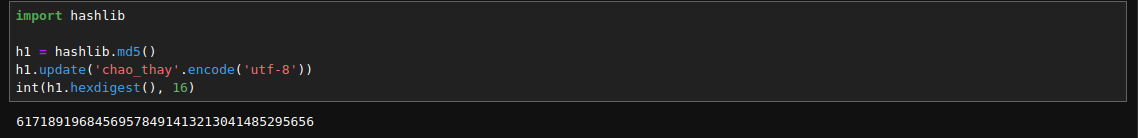
B1: Nạp tiện ích watermark vào môi trường Jupyter Notebook



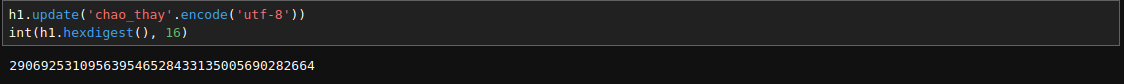
B2: Nạp tên tác giả



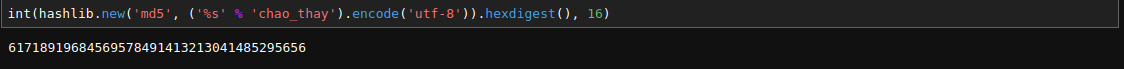
B3: Sử dụng thư viện **hashlib** để tính giá trị hash MD5 của chuỗi 'chao\_thay'



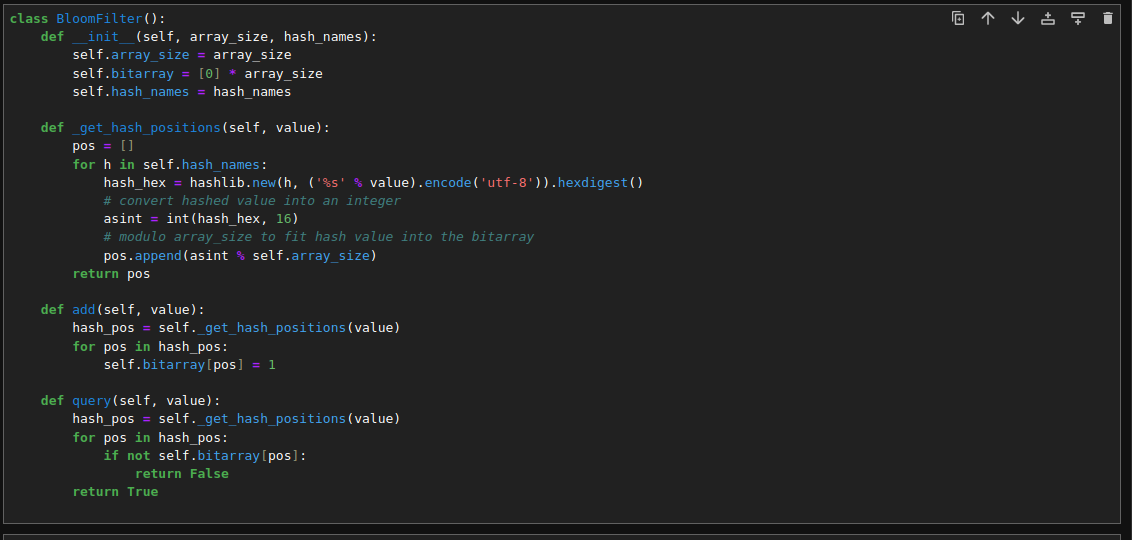
B4: Chuyển đổi chuỗi trên thành số nguyên



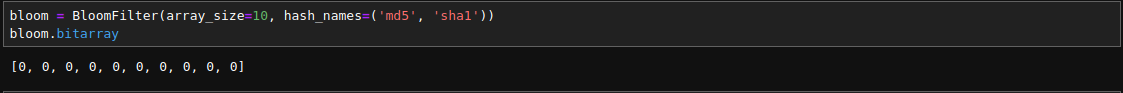
B5: Mã hóa thành dạng utf-8



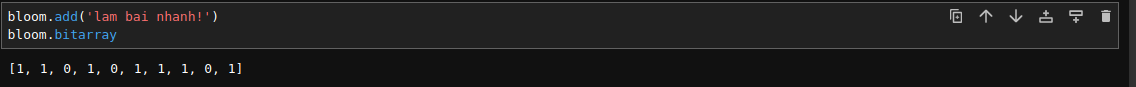
B6: Cài đặt lớp BloomFilter



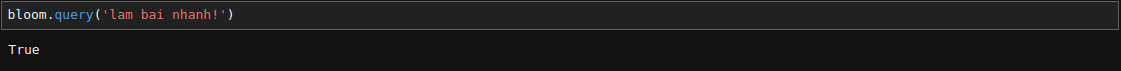
B7: sử dụng một lớp BloomFilter và tạo một đối tượng bloom



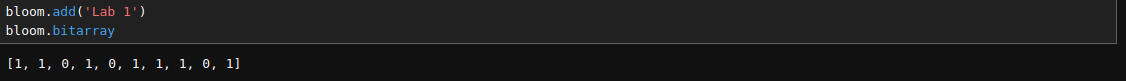
B8: Add chuỗi “lam bai nhanh”



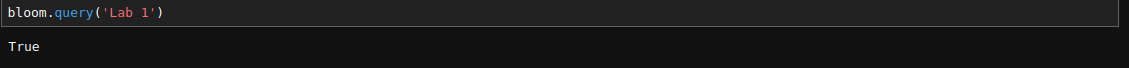
B9: Truy vấn chuỗi vừa add



B10: Add chuỗi “Lab 1”



B11: Truy vấn



B12: Truy vấn chuỗi chưa add

