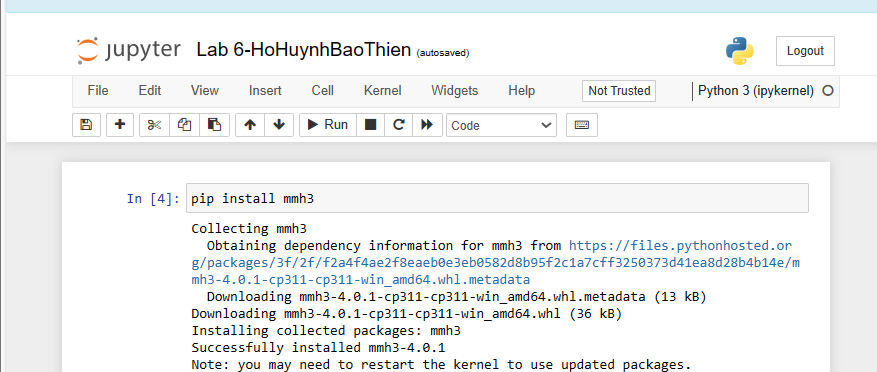
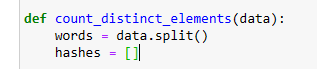
**Cài đặt thư viện mmh3**



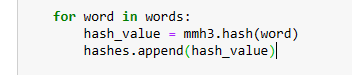
**Gọi thư viện mmh3 (tính toán giá trị băm-hash value) và math (hàm và phương thức)**



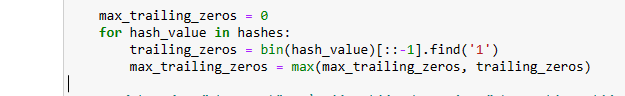
**Tách các từ trong đoạn văn bản thành danh sách các từ riêng biệt và khởi tạo một danh sách các giá trị hash**



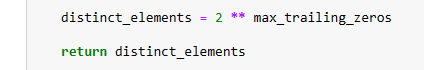
**Tính toán giá trị hash cho mỗi từ và thêm vào danh sách hashes và sử dụng hàm hash mmh3 để tính toán giá trị hash**



**Tìm số lượng bit 0 liên tiếp từ phải sang trái trong giá trị hash lớn nhất**



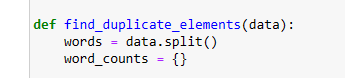
**Tính toán số lượng phần tử riêng biệt dựa trên số lượng bit 0 liên tiếp tìm được**



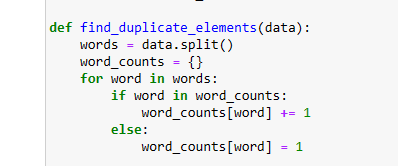
**Tách các từ trong đoạn văn bản thành danh sách các từ riêng biệt**



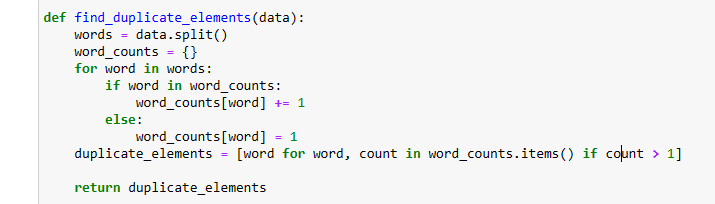
**Tạo một từ điển để lưu trữ số lần xuất hiện của mỗi từ**



**Đếm số lần xuất hiện của mỗi từ**



**Lọc ra các từ có số lần xuất hiện lớn hơn 1**

**Đoạn văn bản đầu vào**

ab23 Big data is a field that treats ways to analyze from or otherwise deal with data Current usage of the term big data tends to predictive analytics user behavior, analytics or advanced data analytics Data sets grow rapidly in part because they are increasingly Internet of things devices such as mobile devices. Xz33. Small Small Small Small Small Small Small Small Small Small Small Small Small Small Small Small



**Đếm số lượng phần tử riêng biệt trong đoạn văn bản**



**Tìm các phần tử lặp lại trong đoạn văn bản**



**Hiển thị số lượng phần tử riêng biệt**



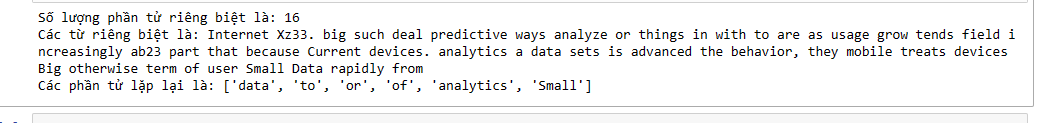
**Hiển thị các từ riêng biệt**

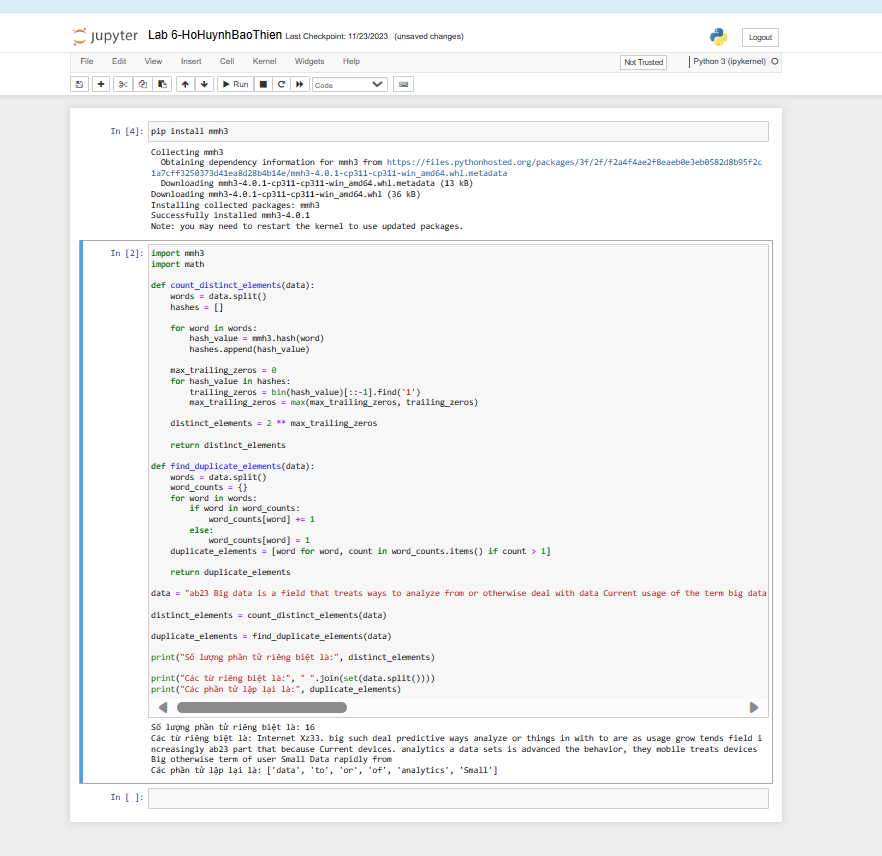
****

**Hiển thị các phần tử lặp lại**



**Kết Quả:**

****

****