Báo cáo bài tập môn Nguyên lý và phương pháp lập trình  
CS111.K21.KHTN  
Giảng viên: Trịnh Quốc Sơn

Thành viên nhóm: Lưu Hoàng Sơn 18521348

Nguyễn Trần Trung 18521555

Bài toán : Đếm số từ có tồn tại trong một danh sách các từ(7700 từ) và một đoạn văn input

1.Ý tưởng

1.1 Ngôn ngữ sử dụng: Python

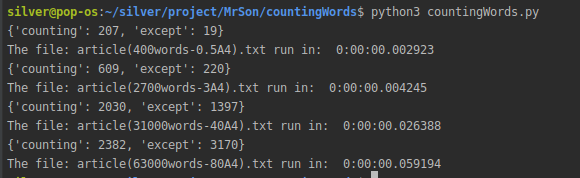
1.2 Các bước thực hiện

* Từ file 7700 từ cho sẵn (syllables) tạo thành một dict với key là các từ và giá trị là True.
* Lí do sử dụng từ dict để tra cứu là bởi vì dict dựa trên bảng hash có tốc độ tìm kiếm nhanh
* Từ file input(test case), tạo list chứa tất cả các từ đã được viết hoa (bỏ qua kí tự đặc biệt)
* Duyệt list tìm kiếm trong dict đã được tạo ở trên (bảng hash) với key là giá trị của từ trong list nếu giá trị trả về là True (nghĩa là có từ đó) thì tăng biến đếm lên 1, False thì bỏ qua, nếu key không tồn tại trong từ điển tăng số lượng từ không tồn tại trong dict lên 1

1.3 Mức độ test:

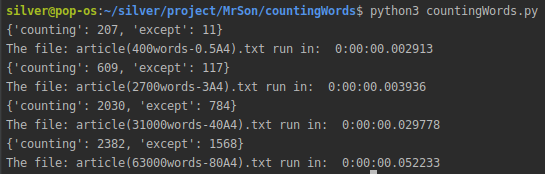
* ~ 400 từ : 0.5 trang A4
* ~ 2700 từ : 3 trang A4
* ~31000 từ : 40 trang A4
* ~63000 từ : 80 trang A4

2. Kết quả



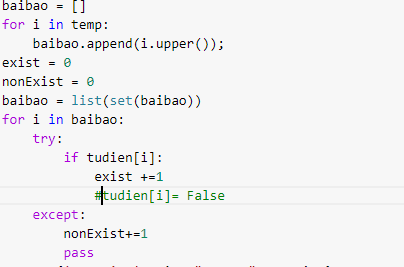
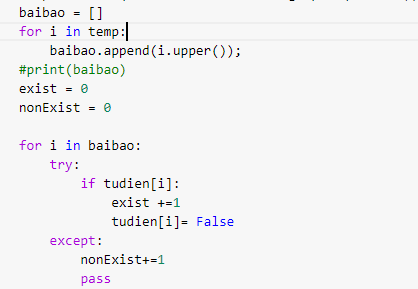
2.1. Cải thiện

Thay vì phải duyệt tất cả các từ có trong input ( sẽ có những từ trùng nhau) ta sẽ loại bỏ những từ trùng và không cần từ điển đánh dấu lại. Vậy trong vòng lặp sẽ tiết kiệm được chi phí tính toán cũng như bỏ được không gian lưu trữ của từ điển.



2.2. Cách cải thiện

Trước Sau:



3. Nhận xét

Code chạy ổn định, khi số lượng dữ liệu đầu vào tăng thì kết quả là thời gian chạy cũng tăng (gần như) tuyến tính

Tuy nhiên số lượng từ không tồn tại trong từ điển đã cho (syllables) cũng ngày càng tăng theo kích thước

* Cần cải thiện file từ điển (tăng kích cỡ từ vựng của file syllables) để có kết quả chính xác hơn ( dữ liệu mẫu chưa đủ)