**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**ĐO ĐỘ TRÙNG KHỚP GIỮA TẤT CẢ CÁC CẶP**

**THÔNG TIN TRONG MỘT THƯ MỤC**

**Môn học: Cấu trúc dữ liệu & giải thuật**

**Giảng viên: Nguyễn Đức Vũ**

**Thành viên nhóm:**

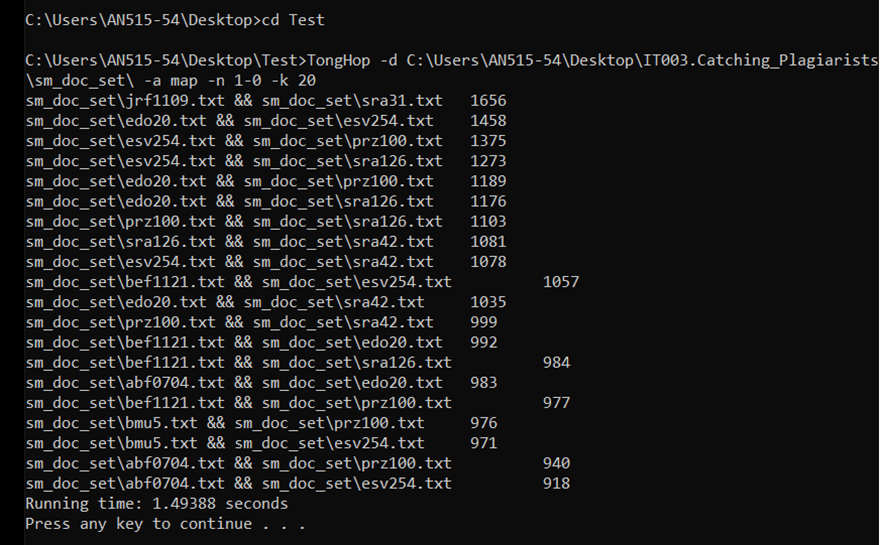
Phạm Xuân Bách \_ 19521233

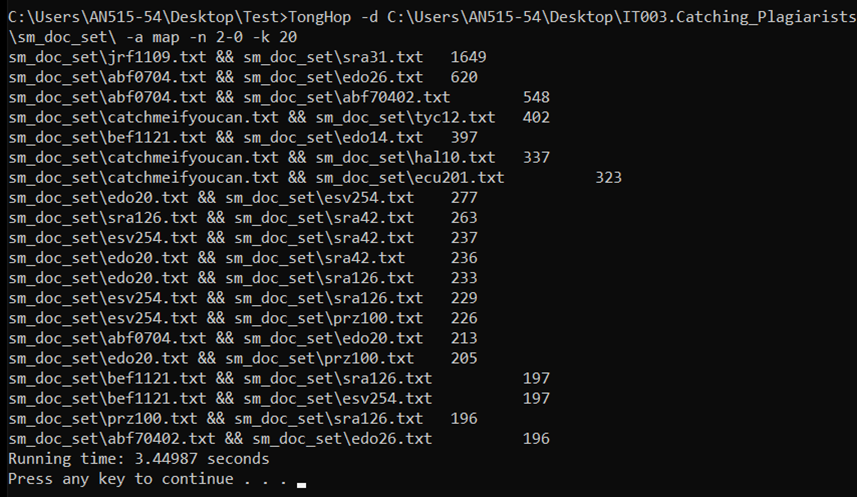
Nguyễn Tiến Đạt \_ 19521345

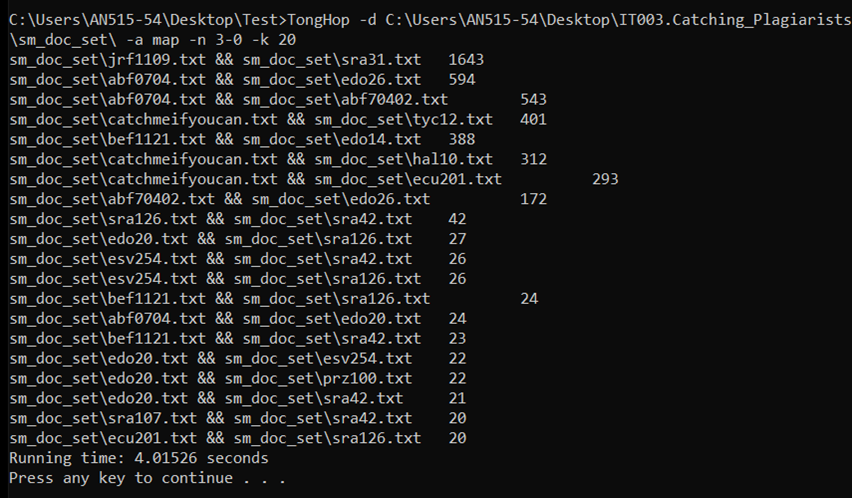
Thành phố Hồ Chí Minh, ngày 26 tháng 7 năm 2020.

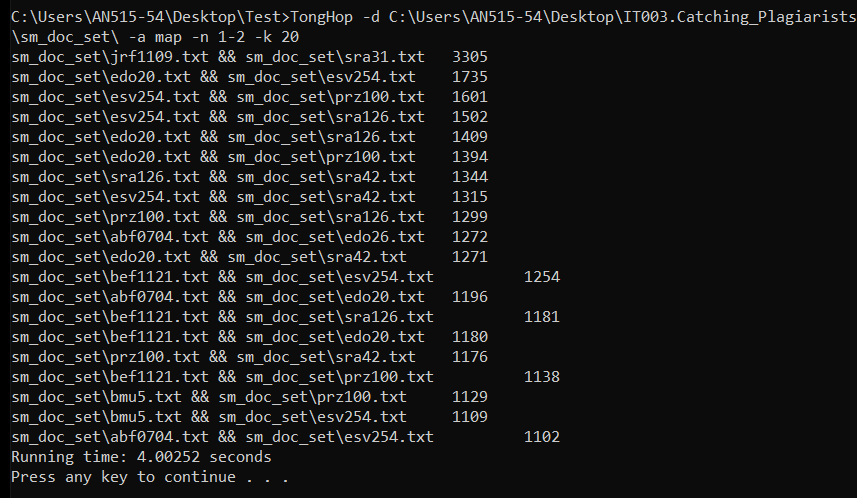
**Nội dung báo cáo:**

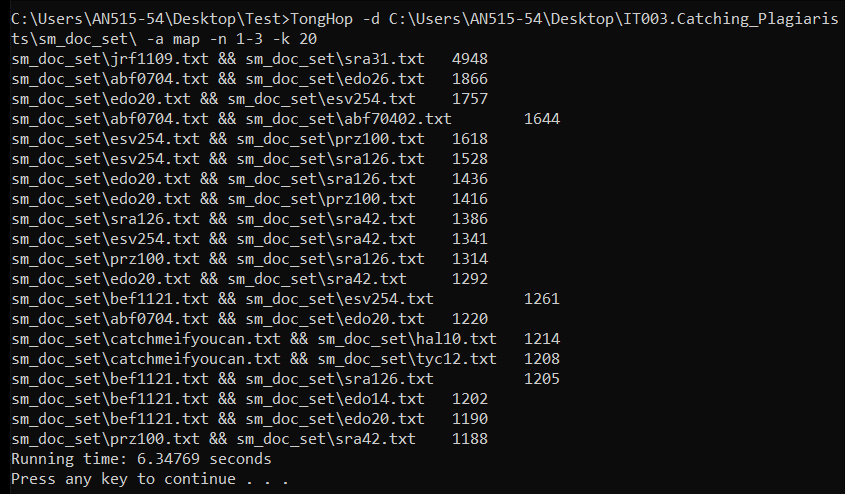
**1. Kết quả 20 trường hợp có độ trùng khớp cao nhất:**

Trường hợp 1: 1-grams

Trường hợp 2: 2-grams

Trường hợp 3: 3-grams

Trường hợp 4: 1-grams, 2-grams

Trường hợp 5: 1-grams, 2-grams, 3-grams

**2. So sánh thời gian (giây) chạy tương đối của chương trình giữa các cấu trúc dữ liệu trên thư mục *sm\_doc\_set***.

Cấu hình máy chạy chương trình:

Hệ điều hành: Window 10 Home Single Language 64-bit.

CPUs: Intel(R) Core(TM) i7-9750H.

RAM: 16GB.

(Đã cắm nguồn).

Trình biên dịch: Code::Blocks 17.12 (Compiler C++14).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tổ hợp đơn vị so khớp | vector | binary tree | bảng băm  (tự cài) | bảng băm  (thư viện map) |
| 1-grams | 3.73732 | 1.25561 | 2.58622 | 1.49388 |
| 2-grams | 6.11032 | 3.16545 | 9.17946 | 3.44987 |
| 3-grams | 6.7293 | 3.15235 | 10.5568 | 4.01526 |
| 1-grams,  2-grams | 9.55571 | 3.19106 | 9.58338 | 4.00252 |
| 1-grams,  2-grams,  3-grams | 15.7223 | 5.17774 | 17.7309 | 6.34769 |

Nhận xét:

* Cấu trúc dữ liệu binary tree (cây nhị phân) có tốc độ xử lý nhanh nhất trong các trường hợp so sánh với các cấu trúc dữ liệu khác.
* Cấu trúc dữ liệu bảng băm (tự cài) có tốc độ xử lý chậm nhất trong các trường hợp so sánh với các cấu trúc dữ liệu khác.
* Cấu trúc dữ liệu bảng băm (thư viện map) có tốc độ xử lý nhanh đứng thứ 2 và cấu trúc dữ liệu vector có tốc độ xử lý nhanh đứng thứ 3 so với các cấu trúc dữ liệu khác.

Đường dẫn tài liệu tham khảo:

<http://www.cplusplus.com/reference/fstream/>

<http://www.cplusplus.com/reference/cstring/strtok/>

<http://www.cplusplus.com/reference/vector/vector/>

<http://www.cplusplus.com/reference/map/map/>

<https://www.tutorialspoint.com/How-can-I-get-the-list-of-files-in-a-directory-using-C-Cplusplus>

<https://www.geeksforgeeks.org/measure-execution-time-function-cpp/>