**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

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

**🙡🕮🙣**

Logo

Description automatically generated

**Search Engine for Truyện Kiều**

**Môn: Truy vấn thông tin Thị giác**

**Giáo viên hướng dẫn:** Võ Hoài Việt

Phạm Minh Hoàng

Phạm Thanh Tùng

**Lớp:** 20TGMT

**Sinh viên thực hiện:** Giang Gia Bảo – 20127446

*Tp Hồ Chí Minh, 16 tháng 06 năm 2023*

**MỤC LỤC**

[I. THÔNG TIN SINH VIÊN 1](#_Toc137890696)

[II. ĐÁNH GIÁ MỨC ĐỘ HOÀN THÀNH 1](#_Toc137890697)

[III. CÁCH CHẠY SOURCE CODE 1](#_Toc137890698)

[IV. ANALYZE AND IMPROVEMENT 2](#_Toc137890699)

[1. Analyze result 2](#_Toc137890700)

[2. Direction for improvement 2](#_Toc137890701)

[V. REFERENCE 3](#_Toc137890702)

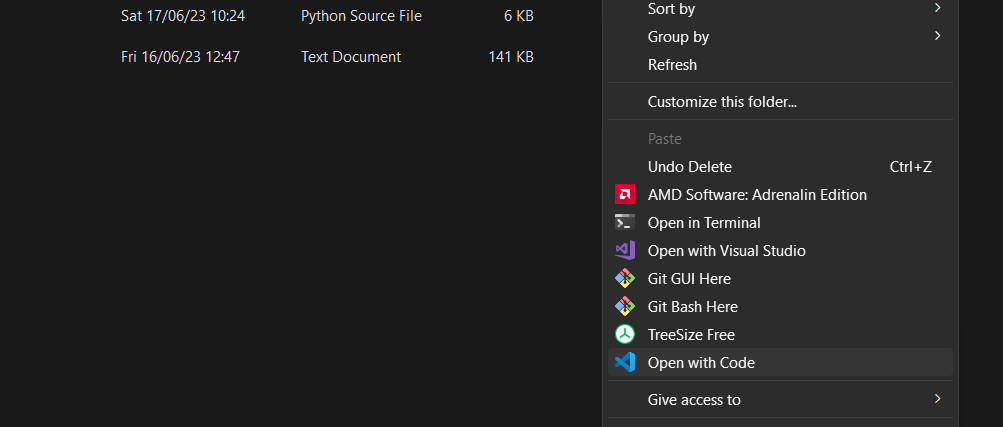
1. THÔNG TIN SINH VIÊN

|  |  |  |  |
| --- | --- | --- | --- |
| **MSSV** | **Họ và tên** | **Lớp** | **Note** |
| 20127446 | Giang Gia Bảo | 20TGMT |  |

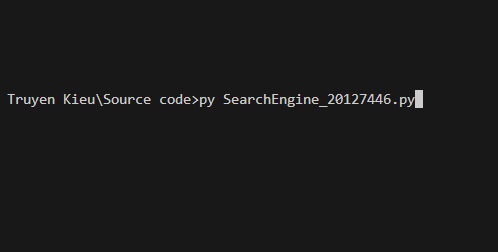
1. ĐÁNH GIÁ MỨC ĐỘ HOÀN THÀNH

|  |  |  |
| --- | --- | --- |
| **STT** | **Yêu cầu** | **Hoàn thành** |
| 1 | Build system | 100% |
| 2 | Evaluate system: precision & recall metrics | 0% |
| 3 | Analyze the result and direction for improvement. | 100% |

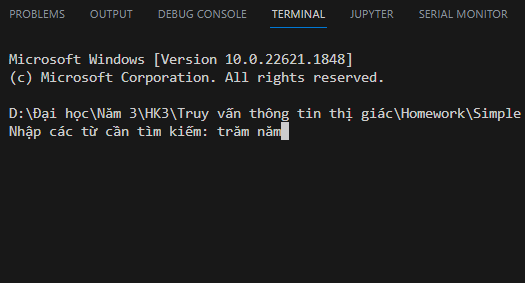
1. CÁCH CHẠY SOURCE CODE
2. **Mở file .py bằng vscode (hoặc các IDE khác)**



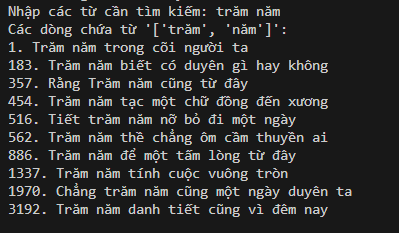
1. **Mở terminal, gõ command để chạy chương trình ‘py SearchEngine\_20127446.py’**

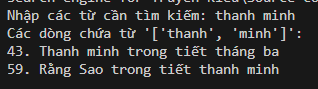


1. **Gõ các từ muốn search (phải có dấu như trong bài thơ gốc)**



1. **Kết quả tìm kiếm**





1. ANALYZE AND IMPROVEMENT
   1. Analyze result

Thông qua kết quả tìm kiếm. Hệ thống có thể tìm từ đơn, hoặc các cụm từ, câu thơ một cách nhanh và chính xác.

Hoạt động bằng cách tìm các dòng chứa từng từ và sau đó lấy các dòng chung chứa tất cả các từ đã tìm kiếm và cho ra kết quả.

Tuy nhiên hệ thống vẫn chưa được phát triển để tìm các từ không dấu, dính nhau.

VD: tram nam, tramnam, …

* 1. Direction for improvement

Hướng phát triển cho hệ thống:

Có thể tìm các từ không dấu như: tram nam, nhung ngay,…

Phát triển mô hình học máy tự sửa lỗi ngữ pháp như sai chính tả, hoặc gợi ý từ gần giống với những từ tìm kiếm không có kết quả. VD: người dùng tìm từ “thãnh thơi”, hệ thống sẽ tự sửa lỗi chính tả và cho ra kết quả của “thảnh thơi”.

1. REFERENCE

<https://github.com/duyet/truyenkieu-word2vec/blob/master/truyen_kieu_data.txt>

<https://helpex.vn/question/cach-tot-nhat-de-loai-bo-cac-dau-trong-chuoi-unicode-python-la-gi-5cb16654ae03f6169c9dcc85>

<https://www.w3schools.com/python/python_dictionaries.asp>