|  |  |
| --- | --- |
| logo | TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP THỰC PHẨM TP.HCM  KHOA CÔNG NGHỆ THÔNG TIN |

**ĐỀ CƯƠNG CHI TIẾT**

**KHÓA LUẬN HỆ ĐẠI HỌC CHÍNH QUY KHÓA 08DHTH**

1. **Tên đề tài:**

**Hệ thống thu thập, phân loại, đánh giá tin tức liên quan đến An ninh quốc gia – Trật tự an toàn xã hội**

1. **Thông tin giảng viên hướng dẫn:**

* Họ tên: Nguyễn Văn Tùng
* Email: tungnv@hufi.edu.vn

1. **Mục tiêu đề tài:**

* Thực hiện xây dựng hệ thống thu thập, phân loại và đánh giá tin tức từ mạng xã hội Facebook. Hệ thống cần thực hiện được các chức năng sau:
* Thu thập tự động những bài viết trên Facebook dựa vào URL
* Phân loại bài viết theo một số chủ đề An ninh quốc gia, An ninh trật tự xã hội.
* Đánh giá bài viết theo một số mức độ như: tích cực – trung bình – tiêu cực
* Ghi nhận lại những bài viết không phù hợp (tiêu cực) và xuất báo cáo.

1. **Yêu cầu**

|  |  |
| --- | --- |
| **Nội dung** | **Điểm (10/10)** |
| * Tìm hiểu cơ chế, phương pháp thu thập tin tức tự động * Trình bày các cơ chế tiền xử lý ngôn ngữ tự nhiên * Tìm hiểu và vận dụng các hàm trong thư viện mã nguồn mở Pyvi, Sklearn và các thư viện liên quan * Tìm hiểu và xây dựng tập mẫu huấn luyện trong việc phân loại, đánh giá tin tức mạng xã hội | 0.5  0.5  0.5  0.5 |
| * Tìm hiểu về ngôn ngữ Python * Tìm hiểu mô hình SAV và ứng dụng vào đề tài. * Tìm hiểu các thuật toán phân loại văn bản * Lập trình công cụ thu thập tin tức từ trang mạng xã hội | 0.5  0.25  0.25  0.5 |
| * Đề xuất phương pháp phân loại, đánh giá tin tức * Có đưa ra các bài toán, hướng giải quyết trong việc phân loại, đánh giá tin tức | 0.25  0.25 |
| * So sánh, đánh giá các thuật toán phân loại khác * Có sử dụng và cải tiến thuật toán trong phân loại, đánh giá tin tức | 0.25  0.25 |
| * Xây dựng mô hình dữ liệu quan hệ * Sơ đồ chức năng hệ thống * Thu thập tin tức từ trang mạng xã hội Facebook * Tạo cơ sở dữ liệu lưu trữ tin tức * Quy trình phân loại, đánh giá tin tức * Phân loại tin tức theo chủ đề chính trị - xã hội * Thống kê những tin tức tiêu cực * Thống kê mức độ tương tác đối với tin tiêu cực * Ghi nhận những tài khoản đăng tin tiêu cực * Đánh giá mức độ ảnh hưởng của tin tức tiêu cực | 0.25  0.25  0.25  0.25  0.25  0.25  0.5  0.5  0.5  0.5 |
| * Nội dung báo cáo: * Trình bày đầy đủ nội dung, định dạng chuẩn (định dạng trang, lề, chữ, đề mục, …) * Chuẩn bị dữ liệu mẫu minh hoạ cho các chức năng đang trình bày (hình kết quả thực nghiệm) * Phong cách báo cáo, slide: * Nội dung powerpoint * Trình bày báo cáo: báo cáo theo quy trình dễ hiểu. | 0.5  0.5  0.5  0.5 |

**4. Môi trường thực hiện:**

* Ngôn ngữ lập trình: Python, C#...
* Môi trường minh hoạ:

**5. Thời gian thực hiện đề tài:**

* 12 tuần (từ ngày 26/10/2020) đến 17/01/2021) đối với khóa luận tốt nghiệp

**6. Thời gian và các công việc trong tuần**:

|  |  |  |
| --- | --- | --- |
| **STT** | **Thời gian** | **Các công việc** |
| 01 | Tuần 1 | * Tìm hiểu cơ chế, phương pháp thu thập tin tức tự động * Trình bày các cơ chế tiền xử lý ngôn ngữ tự nhiên |
| 02 | Tuần 2 | * Tìm hiểu và vận dụng các hàm trong thư viện mã nguồn mở Pyvi, Sklearn và các thư viện liên quan * Tìm hiểu và xây dựng tập mẫu huấn luyện trong việc phân loại, đánh giá tin tức |
| 03 | Tuần 3 | * Tìm hiểu về ngôn ngữ Python * Tìm hiểu mô hình SAV và ứng dụng vào đề tài. |
| 04 | Tuần 4 | * Tìm hiểu các thuật toán phân loại văn bản * Lập trình công cụ thu thập tin tức từ trang mạng xã hội |
| 05 | Tuần 5 | * Đề xuất phương pháp phân loại, đánh giá tin tức * Có đưa ra các bài toán, hướng giải quyết trong việc phân loại, đánh giá tin tức |
| 06 | Tuần 6 | * Xây dựng mô hình dữ liệu quan hệ * Sơ đồ chức năng hệ thống |
| 07 | Tuần 7 | * Thu thập tin tức từ trang mạng xã hội Facebook * Tạo cơ sở dữ liệu lưu trữ tin tức |
| 08 | Tuần 8 | * Quy trình phân loại, đánh giá tin tức |
| 09 | Tuần 9 | * Phân loại tin tức theo chủ đề an ninh xã hội * Thống kê những tin tức tiêu cực * Thống kê mức độ tương tác đối với tin tiêu cực |
| 10 | Tuần 10 | * Ghi nhận những tài khoản đăng tin tiêu cực * Đánh giá mức độ ảnh hưởng của tin tức tiêu cực |
| 11 | Tuần 11 | * Chạy demo hệ thống thu thập, phân loại, đánh giá tin tức * Hoàn thiện slide báo cáo |
| 12 | Tuần 12 | * Hoàn thiện nội dung báo cáo và mô hình demo |

**7. Tài liệu tham khảo**

* https://streetcodevn.com/blog/sav
* <https://topdev.vn/>
* https://viblo.asia/
* https://pypi.org/project/pyvi/
* <https://techtalk.vn/>
* <https://codelearn.io/learning/python-co-ban>
* https://trituenhantao.io/
* https://github.com/deepai-solutions/core\_nlp

**8. Một số yêu cầu khác**

* Số lượng SV thực hiện luận văn: 2 SV/ 1 luận văn
* Có kỹ năng đọc tài liệu Tiếng Anh
* Làm việc nghiêm túc, kỷ luật, chăm chỉ.

**9. Thông tin GVHD**

* Họ tên: Nguyễn Văn Tùng
* Email: [tungnv@hufi.edu.vn](mailto:tungnv@hufi.edu.vn)
* SĐT: 0977977408

Tp.HCM, ngày tháng năm 2020

*Trưởng bộ môn* *Giáo viên hướng dẫn*

(ký và ghi rõ họ tên) (ký và ghi rõ họ tên)