Tên đề tài: Xây dựng luồng thu thập dữ liệu TỪ WEBSITE BÁN KHOÁ HỌC

Nguồn dữ liệu: https://unica.vn/

Trường Đại Học Công Nghiệp TP.HCM

Phân tích định lượng

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mã số sinh viên | Họ và tên | Phân công | Tỉ lệ đóng góp (%) | Điểm số (giảng viên) |
| 22695761 | Nguyễn Thành Đạt | * Crawl dữ liệu(công nghệ, thiết kế xây dựng) * Kiểm soát và sửa lỗi mã nguồn * Xử lý và làm sạch dữ liệu | 40% |  |
| 22665311 | Thái Văn Sơn | * Viết chương trình crawl dữ liệu (unica\_crawler) * Crawl dữ liệu(marketing) | 13% |  |
| 22691511 | Đỗ Viết Thanh Bằng | * Crawl dữ liệu (ngoại ngữ) * Gộp dữ liệu đã crawl (loại bỏ trùng lặp) * Đẩy dữ liệu full lên mongodb | 11% |  |
| 22693501 | Trần Tuấn Sang | * Crawl dữ liệu(Tài chính kế toán) * Thiết kê mô hình thực thể | 13% |  |
| 22696961 | Mai Hoàng Tuấn | * Crawl dữ liệu(sức khoẻ-làm đẹp) * Tạo bảng và đưa dữ liệu đã làm sạch vào postgres | 12% |  |
| 22690271 | Đỗ Tiến Đạt | * Crawl dữ liệu(tin học văn phòng) * Triển khai hệ thống lên trên máy ảo ubuntu server dựa trên docker test. | 11% |  |

Trello: https://trello.com/invite/b/6707e985a3e2c825354f3699/ATTI6405a2570e6ed98c25ad2b7d87a123e9BB4A1DAA/unicaproject Ngày thiết lập: 09/09/2024

Github: https://github.com/3002tad/Unica\_Project Ngày thiết lập: 09/09/2024

Sơ đồ thiết kế hệ thống:

