Tên đề tài: Xây dựng luồng thu thập dữ liệu Từ Website

Nguồn dữ liệu: https://goodreads.com/list/show/19.Best\_for\_Book\_Clubs

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) |
|  | (\*) Leader information on first row |  |  |  |
| 22652711 | Võ Tất Thiên | Khảo sát website, quản lý tiến độ dự án | 15% |  |
| 22724461 | Đỗ Công Danh | Crawl Web | 13% |  |
| 22653661 | Võ Nhật Trường | Đưa dữ liệu vào MongoDB | 13.5% |  |
| 21107481 | Võ Nguyễn Hoành Hợp | Xây dựng cơ sở dữ liệu Postgres | 12.5% |  |
| 22633701 | Trần Quang Vinh | Xây dựng các bảng | 8% |  |
| 22653331 | Nguyễn Thành Kiệt | Làm sạch dữ liệu bằng pyspark, Đưa vào Postgres, Demo máy ảo | 30% |  |
| 22639591 | Nguyễn Đình Huy Hùng | Xây dựng ERD | 8% |  |

Trello: <https://trello.com/b/0FpryE2G/crawldata> Ngày thiết lập: 25/08/2024

Github: <https://github.com/VoTatThien/CrawlWeb_Nhom> Ngày thiết lập: 23/08/2024

Sơ đồ thiết kế hệ thống

A diagram of a diagram

Description automatically generated

A diagram of a diagram

Description automatically generated

A screenshot of a computer

Description automatically generated