

NHẬP MÔN CÔNG NGHỆ PHẦN MỀM

TÀI LIỆU Phân tích & Thiết Kế

Giảng viên: Trần Văn Quý

Sinh viên: Lê Ngọc Huân

Mã số sinh viên: 22880226



Bộ môn Công nghệ phần mềm
Khoa Công nghệ thông tin
Đại học Khoa học tự nhiên TP HCM

MỤC LỤC

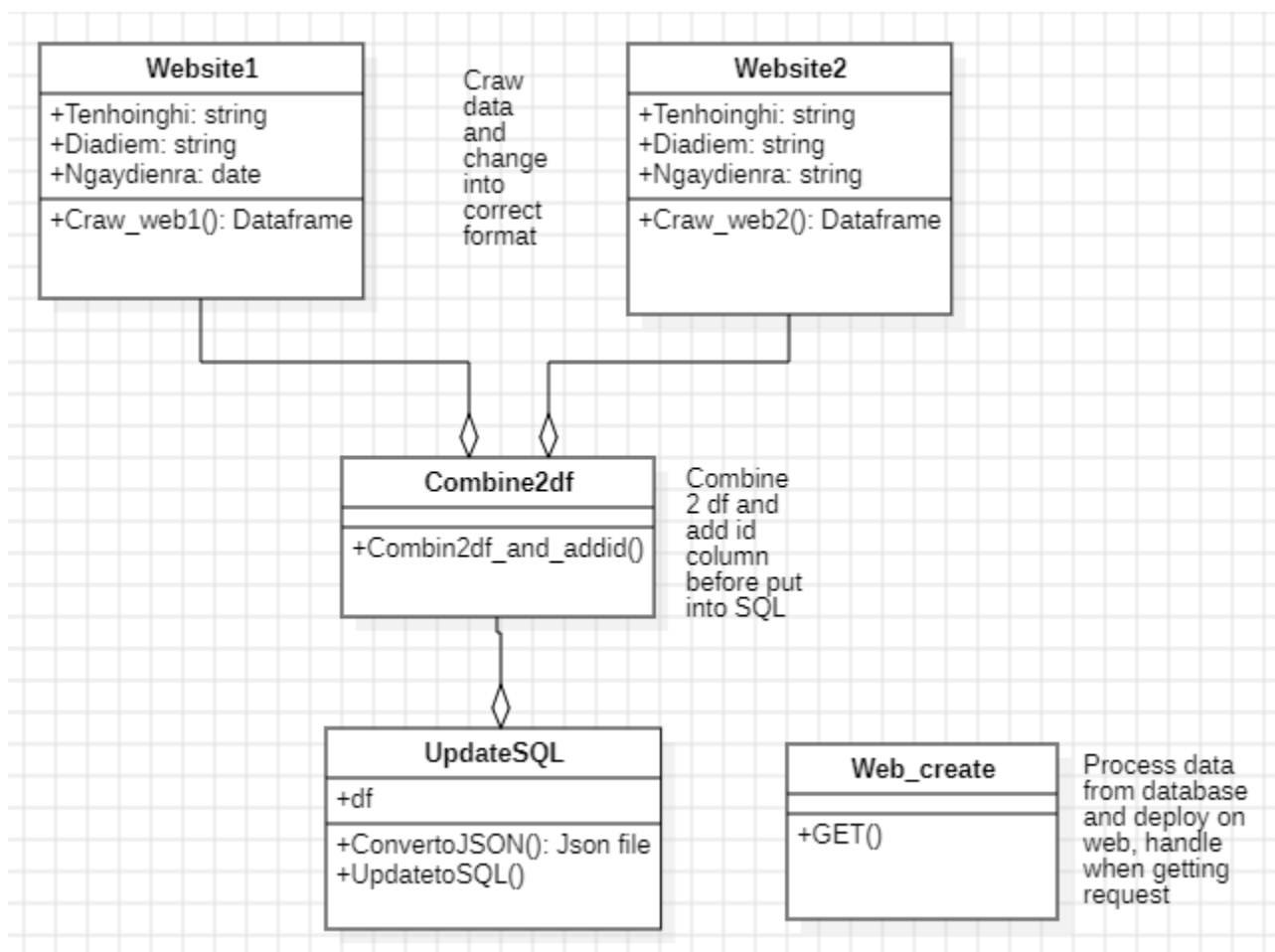
1	Bảng đánh giá thành viên	2
2	Sơ đồ lớp(Class Diagram).....	3
3	Thiết kế kiến trúc	4
4	Thiết kế dữ liệu.....	5
4.1	Đặc tả dữ liệu.....	Error! Bookmark not defined.

TÀI LIỆU THIẾT KẾ

1 Bảng đánh giá thành viên

MSSV	Họ Tên	% đóng góp (tối đa 100%)	Chữ ký
22880226	Lê Ngọc Huân	100%	Huân

2 Sơ đồ lớp(Class Diagram)



3 Thiết kế kiến trúc

Tổng Quan Kiến Trúc Hệ Thống

Hệ thống có thể được chia thành bốn phần chính:

Crawling Engine: Thu thập dữ liệu từ các trang web.

Data Processing Module: Xử lý và chuẩn hóa dữ liệu thu thập được.

Database: Lưu trữ dữ liệu đã được xử lý.

Web Application: Giao diện người dùng để tương tác với dữ liệu.

Crawling Engine

Process: Sử dụng thư viện như BeautifulSoup trong Python để truy xuất và thu thập dữ liệu từ các trang web. Các crawler được thiết kế để tránh thu thập dữ liệu không được phép.

Data Processing Module

Input: Dữ liệu thô từ Crawling Engine.

Process:

Data Cleaning: Loại bỏ dữ liệu hỏng hoặc không cần thiết.

Data Transformation: Chuyển đổi dữ liệu vào định dạng cần thiết (ví dụ: định dạng ngày tháng, chuẩn hóa tên).

Output: Dữ liệu đã được chuẩn hóa được gửi tới Database.

Database

Technology: Sử dụng Microsoft SQL Server 12

Structure: Thiết kế sơ đồ cơ sở dữ liệu để hỗ trợ truy vấn hiệu quả và lưu trữ an toàn.

Web Application

Frontend: Sử dụng HTML, hiển thị dữ liệu, chức năng tìm kiếm, bao gồm các bảng dữ liệu và kết quả tìm kiếm

Backend: Flask, một micro-framework Python, Xử lý các yêu cầu từ frontend, thực hiện truy vấn đến cơ sở dữ liệu và trả dữ liệu lại cho người dùng

4 Thiết kế dữ liệu

4.1 Đặc tả dữ liệu

Dữ liệu hiện được lưu vào 1 bảng duy nhất do được crawl về không nhiều. Sau đó dữ liệu được dùng để hiển thị trên web tương tác.

Conferences			
	Column Name	Data Type	Allow Nulls
🔑	ID_Conference	char(10)	<input type="checkbox"/>
	Name_Conference	nvarchar(200)	<input checked="" type="checkbox"/>
	Place	nvarchar(100)	<input checked="" type="checkbox"/>
	Time	date	<input checked="" type="checkbox"/>
			<input type="checkbox"/>