ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Xây dựng module phân loại dữ liệu, gợi ý, tìm kiếm và bảo mật trong hệ thống học và thi online worksheetzone

LÊ ĐĂNG HOÀNG ĐẠT
dat.ldh183882@sis.hust.edu.vn

Ngành Công nghệ thông tin

Giảng viên hướng dẫn: TS. Nguyễn Thanh Hùng _____

Chữ kí GVHD

Khoa: Công nghệ thông tin

Trường: Công nghệ thông tin và Truyền thông

HÀ NỘI, 02/2023

LỜI CẢM ƠN

Đầu tiên, em muốn cảm ơn chính bản thân mình vì những cố gắng, nỗ lực tới cùng để không phải từ bỏ cho dù gặp nhiều khó khăn trong giai đoạn thực hiện đồ án và cũng cảm ơn vì đã hoàn thành được con đường tại Bách Khoa - một con đường tuy không hề bằng phẳng, dẫu có lúc phân vân vì lựa chọn của bản thân nhưng cuối cùng em đã đúng với lựa chọn của mình.

Em xin cảm ơn thầy Nguyễn Thanh Hùng, đối với em thầy không chỉ là người thầy mà thầy còn là một người hướng dẫn tuyệt vời trong quãng đời sinh viên của em. Nếu không có thầy giúp đỡ trong những năm gắn bó tại Bách Khoa, cuộc đời em có thể đã rẽ hướng sang một nhánh khác thay vì trở thành một kỹ sư công nghệ phần mềm như định hướng hiện tại. Những lời cảm ơn có lẽ là không đủ để có thể nói hết lòng biết ơn nhưng từ trong lòng, em xin chân thành cảm ơn thầy!

Em thật sự muốn gửi lời cảm ơn đến công ty cổ phần ABC cùng các đồng nghiệp đã hỗ trợ hết mình trong quá trình em làm việc tại đây.

Ngoài ra, em cũng muốn gửi lời cảm ơn chân thành đến với đại gia đình Liên chi Đoàn- Liên chi Hội Trường Công nghệ thông tin và Truyền thông đã cho em một môi trường gắn bó với những người anh, người bạn, người em đáng nhớ. Cảm ơn anh Tài Phan, anh Thái Bảo, anh Đình Dương, Mạnh Trường, Tuấn Sơn, em Quốc Trung, rất vui vì chúng ta đã có những khoảng thời gian đáng nhớ, phấn đấu để cùng nhau đi lên.

Cảm ơn Mạnh Hùng, Phương Anh, Thuỷ Tiên và đặc biệt Bích Phượng vì những kỷ niệm rất đẹp trong suốt quãng đời đại học của chúng ta.

Và cuối cùng, cảm ơn sự đồng hành của bố mẹ, của những người thầy dạy dỗ từ những ngày đầu tiên để rèn rũa, uốn nắn bản thân em trở thành một con người tốt, lương thiện và có ích cho xã hội.

Cảm ơn Bách Khoa, cảm ơn vì những năm tháng thật sự rực cháy, sống hết mình với thứ gọi là tuổi trẻ.

TÓM TẮT NỘI DUNG ĐỒ ÁN

Ngày nay với sự phát triển của cuộc cách mạng 4.0, việc chuyển dịch các hoạt động từ trực tiếp sang trực tuyến là nhu cầu hết sức thiết thực để có thể mở rộng quy mô cũng như lan tỏa thêm sức ảnh hưởng mà các hoạt động này mang lại và việc học và thi cũng không nằm ngoài những hoạt động đó. Cùng thời điểm bùng nổ về sự tiến bộ vượt bậc của khoa học kỹ thuật, dịch bệnh Covid 19 đã có những ảnh hưởng không hề nhỏ trong quá trình chuyển dịch các hoạt động giáo dục qua hình thức trực tuyến.

Việc xây dựng hệ thống học và thi trực tuyến nhằm đơn giản hóa việc học của học sinh trong thời gian giãn cách xã hội là một giải pháp tốt nhằm giải quyết vấn đề này. Trong đó, việc phân loại dữ liệu trong ngân hàng đề thi, cùng với đó là gợi ý tìm kiếm những đề thi có liên quan tới người sử dụng nhằm giúp giảm thiểu thời gian tìm kiếm những đề thi theo nhu cầu là một mô đun rất cần thiết. Bên cạnh đó, việc bảo mật cho hệ thống học và thi trực tuyến cũng là một vấn đề cần được chú trọng.

Từ tình hình thực tế, hướng tiếp cận của em là nghiên cứu một hệ thống phân loại dữ liệu, xây dựng thuật toán gợi ý, tìm kiếm các đề thi cùng với đó là tích hợp một thư viện đảm bảo các giải pháp bảo mật cơ bản nhất cho hệ thống học và thi trực tuyến. Qua quá trình nghiên cứu và phát triển, hệ thống đã đáp ứng được những yêu cầu cơ bản nhất về mặt bảo mật như phòng chống các lỗi bảo mật cơ bản nhất với một hệ thống website như chống giả mạo người dùng, mã hóa những dữ liệu cần thiết, đóng dấu bản quyền với chữ ký ảnh cùng với đó là một hệ thống phân loại dữ liệu, gợi ý tìm kiếm những đề thi liên quan đến dữ liệu tới từ người sử dụng.

ABSTRACT

Today, with the development of the Fourth Industrial Revolution, the shift from face-to-face to online activities has become a practical necessity to expand the scale and increase the reach of activities. Learning and exams are no exception to this trend. At the same time, the rapid progress of science and technology, coupled with the impact of the COVID-19 pandemic, has accelerated the transition of educational activities to an online format.

Developing an online learning and testing system to facilitate students' education during social distancing is an effective solution to this problem. Specifically, organizing exam questions into a searchable database, along with providing relevant question suggestions, can significantly reduce the time required to locate relevant questions. Furthermore, ensuring the security of the online learning and exam system is critical.

Given the current situation, my approach involves researching a data classification system, designing a suggestion algorithm for question searches, and integrating a library to provide security solutions. Security is the most basic requirement for an online learning and exam system. Through the research and development process, the system has met the most basic security requirements, such as preventing common security errors in website systems (e.g., user anti-forgery and data encryption) and incorporating copyright signatures with photo identification. Additionally, the system includes a data classification system and provides question suggestions to enhance the user experience.

MỤC LỤC

DANH MỤC HÌNH VẼ

DANH MỤC BẢNG BIỂU

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI

1.1 Đặt vấn đề

Thời điểm hiện nay, những hệ thống website hỗ trợ việc học và thi trực tuyến đã xuất hiện rất nhiều, nhưng một hệ thống có thể giúp các giáo viên, phụ huynh cũng như học sinh tạo ra các đề thi cũng như tìm kiếm các tài liệu để luyện tập với đa dạng các thể loại bài tập thì chưa có hệ thống nào thật sự đáp ứng đủ tốt. Trong quá trình xây dựng và phát triển hệ thống học và thi trực tuyến Worksheetzone, em nhận ra một số vấn đề cần phải giải quyết để cải thiện chất lượng hệ thống như:

- Phân loại các dữ liệu hiện có để có thể gợi ý, tìm kiếm dễ dàng hơn đối với dữ liệu, mục đích của từng cá nhân
- Xây dựng các giải pháp bảo mật để nâng cao chất lượng, phòng chống các cuộc tấn công tới hệ thống, gây ra những bất lợi không đáng có cho hệ thống

Phân loại các dữ liệu của đa phần các hệ thống hiện tại chỉ dựa vào một số trường cụ thể như tên đề thi, mã đề thi,... nhưng chưa thật sự có thể gợi ý các đề thi cũng như nguồn tài liệu dựa vào dữ liệu người dùng.

Lí do em lựa chọn xây dựng thêm một hệ thống bảo mật là bởi vì trong quá trình nghiên cứu và phát triển hệ thống phân loại dữ liệu, gợi ý và tìm kiếm đề thi cho hệ thống Worksheetzone, em đồng thời phát hiện ra những yếu điểm về mặt bảo mật của hệ thống này. Cùng với đó, các phương án bảo mật hiện nay chủ yếu được xây dựng theo từng trường hợp cụ thể và thông thường sẽ chỉ được sử dụng đến như một bản vá lỗi khi hệ thống đã bị tấn công hoặc nguồn tài nguyên đã bị đánh cắp. Nhìn nhận thực tế cho thấy rằng, chưa có một gói giải pháp bảo mật nào cung cấp đầy đủ những cách thức bảo vệ cho hệ thống website cùng với đó là việc gặp khó khăn khi phải cài đặt từng giải pháp cho từng phương án bảo mật cụ thể.

Từ các vấn đề thực tiễn kể trên, trong đồ án tốt nghiệp lần này, em xin đưa ra một hệ thống phân loại dữ liệu, gợi ý, tìm kiếm và bảo mật cho hệ thống học và thi trực tuyến Worksheetzone nhằm giải quyết bài toán gợi ý, tìm kiếm các đề thi, nguồn tài liệu dựa theo dữ liệu người dùng qua việc phân loại dữ liệu cùng với đó là tích hợp một gói giải pháp bảo mật nhằm ngăn chặn những lỗ hổng cơ bản nhất đối với hệ thống website Worksheetzone.

1.2 Mục tiêu và phạm vi đề tài

Đối với hệ thống học và thi trực tuyến Worksheetzone hiện tại, công cụ tìm kiếm chủ yếu là lọc chính xác theo tên của các đề thi, chưa có công cụ để có thể lọc các đề thi và tài liệu theo các chủ đề cũng như các khối lớp, điều này khiến cho việc tìm