

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM TP. HỒ CHÍ MINH



BÁO CÁO
ĐỀ TÀI: XÂY DỰNG CHƯƠNG TRÌNH TRA CỨU
KIẾN THỨC MÔN LẬP TRÌNH CƠ BẢN

MÔN CÁC HỆ CƠ SỞ TRI THỨC

Sinh viên:

Lâm Hoàng Khánh – 44.01.104.112

Trần Gia Phát – 44.01.104.165

Lâm Phát Tài – 44.01.104.188

Thành phố Hồ Chí Minh, ngày 18 tháng 12, năm 2021

Mục Lục

Chương 1.	GIỚI THIỆU	1
1.1.	Giới thiệu:.....	1
1.2.	Mục tiêu:	1
1.3.	Giới hạn:.....	1
Chương 2.	THU THẬP TRI THỨC	2
Chương 3.	TỔ CHỨC TRI THỨC	3
Chương 4.	TRUY VẤN TRI THỨC	3
4.1.	Phương pháp truy vấn	3
4.2.	Quá trình truy vấn tri thức.....	4
Chương 5.	GIAO DIỆN	5

Chương 1. GIỚI THIỆU

1.1. Giới thiệu:

Lập trình cơ bản là một môn học rất quan trọng đối với sinh viên ngành Công Nghệ Thông Tin, nó giúp sinh viên hiểu được các khái niệm cũng như tư duy logic khi lập trình. Mặc dù quan trọng nhưng thời gian giảng dạy và thực hành trên lớp có giới hạn nên việc các bạn sinh viên tự trao đổi thêm kiến thức là một nhu cầu cần thiết.

Nhưng kiến thức trong lĩnh vực lập trình thì tương đối rộng và mỗi tài liệu có cách diễn đạt khác nhau dẫn đến việc các bạn sẽ gặp không ít khó khăn trong quá trình tự học của mình. Hiểu được vấn đề đó, trong môn học các hệ cơ sở tri thức này, nhóm chúng em đề xuất xây dựng một chương trình hỗ trợ tra cứu kiến thức môn lập trình cơ bản dựa trên hệ cơ sở tri thức được thu thập.

Nhóm chúng em hy vọng hệ thống sẽ là một công cụ hữu ích giúp các bạn có nhu cầu tự học lập trình cơ bản có thể tra cứu được những thông tin cần thiết một cách nhanh chóng và chính xác. Bên cạnh đó, việc xây dựng hệ thống cũng giúp nhóm ôn tập lại những kiến thức cũ cũng như trao đổi thêm kĩ năng trong xử lí ngôn ngữ tự nhiên Tiếng Việt và xây dựng mô hình máy học.

1.2. Mục tiêu:

Những mục tiêu mà nhóm muốn đạt được trong đề tài này bao gồm:

- Thu thập những kiến thức cơ bản nhưng quan trọng trong môn lập trình cơ bản C++. Bao gồm biến, toán tử, vòng lặp for, câu lệnh rẽ nhánh, hàm, chuỗi.
- Từ những kiến thức đã thu thập tiến hành tổng hợp thành một kho ngữ liệu ở để phục vụ cho việc xây dựng hệ cơ sở tri thức.
- Với cơ sở tri thức, nhóm xây dựng một hệ thống tra cứu kiến thức cơ bản về ngôn ngữ C++.
- Hệ thống bao gồm mô hình máy học xử lí ngôn ngữ tự nhiên Tiếng Việt để dự đoán được ý định và ngữ cảnh trong câu nói của người dùng. Từ ý định và ngữ cảnh kết hợp với thuật ngữ lập trình trong câu hỏi, hệ thống có thể đưa ra câu trả lời phù hợp với yêu cầu của người dùng.
- Xây dựng giao diện web cho hệ thống để người dùng có thể dễ dàng tương tác.

1.3. Giới hạn:

Kiến thức lập trình thì tương đối rộng lớn, vì vậy nhóm chúng em xin giới hạn đề tài

- Về kiến thức:

Chủ yếu trả lời những câu hỏi ở dạng khái niệm.

- Về ứng dụng:

Cung cấp một vài hàm xử lý cơ bản để người dùng có thể áp dụng nhanh giúp người dùng tránh việc học những khái niệm khô khan.

Chương 2. THU THẬP TRI THỨC

Hệ thống lại kiến thức lập trình C++ cơ bản, chúng tôi nhận thấy tri thức cần bao gồm các chương sau:

1. Thông tin giới thiệu về ngôn ngữ lập trình C++
2. Nhập/ Xuất dữ liệu
3. Cấu trúc điều kiện (if-else; switch-case)
4. Cấu trúc lặp (for, while,...)
5. Hàm
6. Mảng, Ký tự, Chuỗi
7. Cấu trúc (struct)

Chúng tôi tiến hành thu thập các văn bản ngôn ngữ tự nhiên về đề tài lập trình C++ để tổng hợp thành một kho ngữ liệu lập trình C++. Nguồn thu thập từ các trang website uy tín và giáo trình dưới đây:

1. Giáo trình Bài giảng C++
(Phạm Hồng Thái - TRƯỜNG ĐẠI HỌC CÔNG NGHỆ ĐẠI HỌC QUỐC GIA HÀ NỘI)
2. Giáo trình Lập trình C++
(Trần Hữu Quốc Thư - TRƯỜNG ĐẠI HỌC SƯ PHẠM TP.HCM)
3. <https://www.cplusplus.com>
4. <https://viettuts.vn/lap-trinh-cpp>

Dữ liệu được các thành viên trong nhóm thu thập cách thủ công, được lưu dưới dạng txt (text document). Do giới hạn về thời gian, nên nhóm đã bước đầu thu thập được các định nghĩa, cú pháp chương đề ra.

Bên cạnh các dữ liệu tri thức về lập trình C++ cơ bản, nhóm còn thu thập thêm dữ liệu về các câu hỏi lập trình nhằm phục vụ cho việc huấn luyện môn hình phân loại ý định của người dùng (User Intent Classification) và phân loại ngữ cảnh (Context Classification). Phần lớn các câu hỏi được thu thập từ các thắc mắc của sinh viên Khoa Công nghệ Thông tin trường Đại học Sư phạm TP.HCM.

Dữ liệu được tổ chức thành một tập các bộ gồm 1000 câu hỏi hoặc câu nói cùng với ngữ cảnh và ý định tương ứng của người dùng. Câu hỏi hoặc câu nói chính là dữ liệu đầu vào cho các mô hình máy học. Ngữ cảnh hay ý định tương ứng với mỗi câu sẽ là nhãn cho 2 mô hình máy học (mô hình dự đoán ngữ cảnh và mô hình dự đoán ý định).

Chương 3. TỔ CHỨC TRI THỨC

Các tri thức được tổ chức theo thuật ngữ lập trình, ngữ cảnh và ý định của người dùng theo cấu trúc sau:

“Thuật ngữ lập trình”: {

“Ngữ cảnh”: {

“Dạng câu truy vấn 1” (ý định): [“nội dung”],

“Dạng câu truy vấn 2” (ý định): [“nội dung”],

“Dạng câu truy vấn 3” (ý định): [“nội dung”],

“Nguồn”: [“nội dung”]

}

}

Về ngữ cảnh sẽ bao gồm 2 lớp là các câu hỏi về thông tin hệ thống hoặc thông tin về lập trình nhưng không đề cập cụ thể ngôn ngữ nào (General); các câu hỏi, thắc mắc về lập trình C++.

Về ý định người dùng sẽ bao gồm 8 lớp:

1. Agree
2. Disagree
3. Info
4. Greeting
5. Define
6. Compare
7. Apply
8. Other

Cuối cùng là nguồn sẽ là tên các giáo trình, đường dẫn trang website nơi được trích xuất nội dung thông tin về lập trình đang được đề cập.

Chương 4. TRUY VẤN TRI THỨC

4.1. Phương pháp truy vấn

Từ câu hỏi mà người dùng nhập vào, 2 mô hình máy học sẽ xác định ý định và ngữ cảnh của cuộc đối thoại sau đó câu trả lời được truy vấn từ hệ tri thức đã được tạo ra từ trước.

Có 3 loại kiến thức chính của hệ thống để truy vấn là:

- Define (Định nghĩa của tri thức).

Hệ thống trả về những câu trả lời là định nghĩa, khái niệm của thuật ngữ lập trình trong ngôn ngữ C++.

Ví dụ: *“String trong C++ là gì?”*.

- Apply (Ứng dụng của tri thức).

Hệ thống trả về những câu hỏi mang tính ứng dụng, cách sử dụng và có thể đưa ra những ví dụ cho tri thức đó.

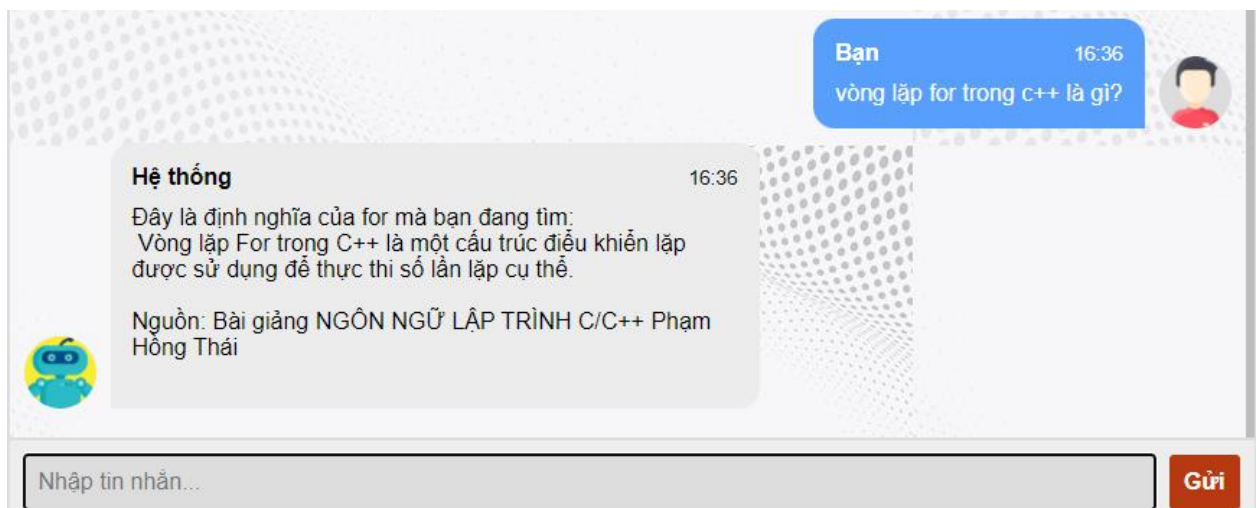
Ví dụ: “*Cách sử dụng String như thế nào?*”.

- Compare (So sánh giữa 2 tri thức).

Trước tiên hệ thống tra cứu sẽ xác định đây có phải là loại câu hỏi so sánh không, nếu có thì tiếp tục xác định 2 miền tri thức cần được so sánh là gì, sau đó hệ thống sẽ trả về những tri thức đã được nhóm chuẩn bị từ trước.

Ví dụ: “*So sánh vòng lặp for và vòng lặp while*”. (Đây là câu hỏi so sánh có chứa 2 thuật ngữ lập trình là “For” và “While”).

4.2. Quá trình truy vấn tri thức



Hình 1: Câu hỏi định nghĩa cần truy vấn.

Ở hình trên, người dùng nhập vào câu hỏi “*Vòng lặp for trong c++ là gì?*”. Trước hết, hệ thống sẽ xác định câu hỏi mang ý định Define và thuộc ngữ cảnh C++. Từ đó, hệ thống sẽ tra cứu trong miền tri thức được lưu trữ và nó trả về kết quả tương ứng (Hình 2).

```

"for": {
  "c++": {
    "define": ["Vòng lặp For trong C++ là một cấu trúc điều khiển lặp được sử dụng để thực
thi số lần lặp cụ thể."],
    "apply": ["Cú pháp vòng lặp for\nfor (khởi_tạo_biến ; check_điều_kiện ; tăng/
giảm_biến) { \n    // Khối lệnh được thực thi\n}\n\nVí dụ vòng lặp for\nfor (int a =
10; a < 20; a++) {\n    cout << 'Gia trị của a: ' << a << endl;\n}"],
    "compare": ["Vòng lặp For trong C++ có cú pháp phức tạp, nhưng ngắn gọn hơn các vòng
lặp while hay do-while khi sử dụng. Vòng lặp for hoàn toàn có thể thay thế vòng lặp
while. Vòng lặp for thường được sử dụng cho các trường hợp biết trước số lần lặp lại
khối công việc."],
    "__source__": ["Bài giảng NGÔN NGỮ LẬP TRÌNH C/C++ Phạm Hồng Thái"]
  }
},

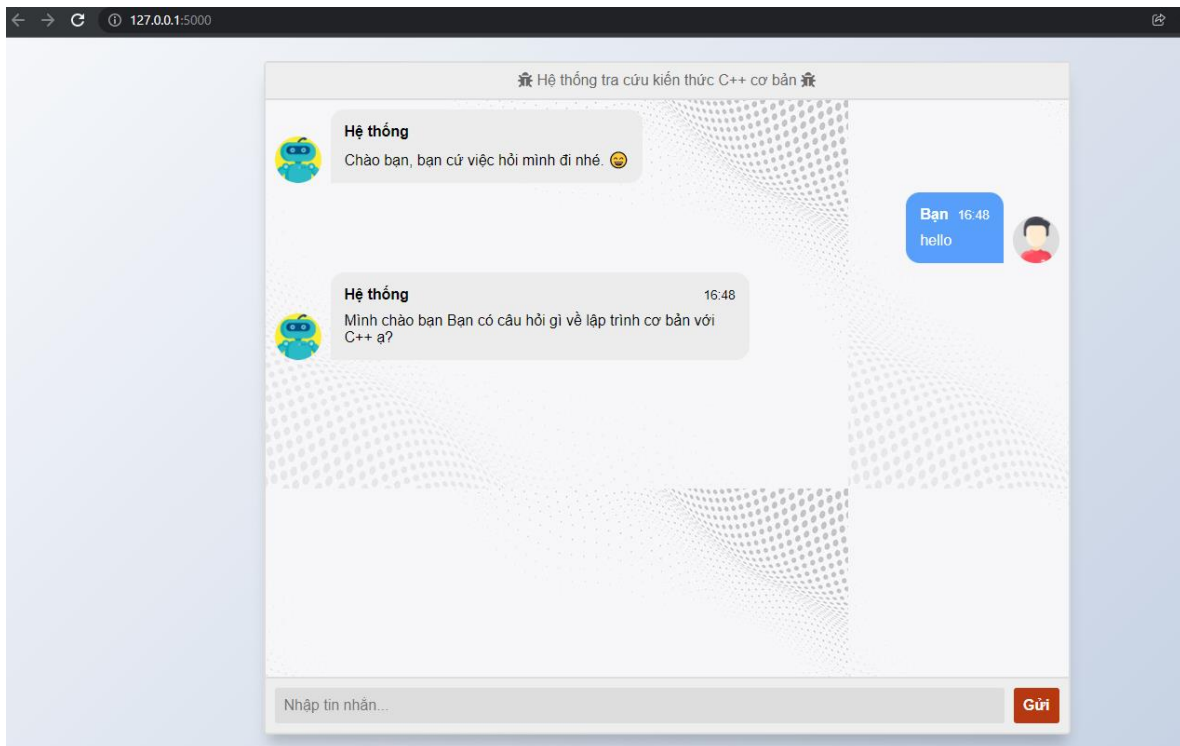
```

Hình 2: Tri thức về thuật ngữ lập trình “For”.

Chương 5. GIAO DIỆN

Về phần giao diện, nhóm sử dụng ngôn ngữ thiết kế đơn giản là HTML và CSS để làm giao diện quen thuộc dễ sử dụng.

Bên cạnh đó, nhóm còn sử dụng thêm thư viện Flask của Python để kết nối khối giao diện đã được thiết kế với server để trả về tri thức tương ứng.



Hình 3: Giao diện hoàn chỉnh của hệ thống.

-Hết-