BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC SƯ PHẠM TPHCM**

**THUYẾT MINH ĐỀ TÀI**

**NGHIÊN CỨU KHOA HỌC CỦA SINH VIÊN**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. **TÊN ĐỀ TÀI:** **CodEbot – Một hệ thống chatbot sử dụng ngôn ngữ Tiếng Việt để giải đáp thắc mắc trong lập trình cơ bản** | | | | | | | | | | | | | | | | | | | | 1. **MÃ SỐ** | | | | | | |
|  | | | | | | |
| 1. **LĨNH VỰC NGHIÊN CỨU** | | | | | | | | | | | | | | | | | | | | 1. **LOẠI HÌNH NGHIÊN CỨU** | | | | | | |
| Tự nhiên | | | Xã hội nhân văn | | Giáo dục | | | Kỹ thuật | | | Nông  Lâm-Ngư | | | | Y dược | | Môi trường | | | Cơ bản | | Ứng dụng | | | Triển khai | |
| ☐ | | | ☐ | | ☒ | | | ☒ | | | ☐ | | | | ☐ | | ☐ | | | ☐ | | ☒ | | | ☐ | |
| 1. **THỜI GIAN THỰC HIỆN** | | | | | | | | | | 06 tháng,  từ tháng 09 năm 2019 đến tháng 04 năm 2020 | | | | | | | | | | | | | | | | |
| 1. **CƠ QUAN CHỦ TRÌ** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Khoa/Bộ môn trực thuộc: Khoa Công Nghệ Thông Tin | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1. **SINH VIÊN THỰC HIỆN ĐỀ TÀI** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Họ và tên: Vương Lê Minh Nguyên Mã số SV: 43.01.104.117  1/ Vương Lê Minh Nguyên - 43.01.104.117 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Địa chỉ NR: | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Điện thoại: | | | | | | | | | | | |  | | | | | | |  | | | | | | | |
| E-mail: vlmnguyen98@gmail.com  **GIÁO VIÊN HƯỚNG DẪN**  Họ và tên: TS. Nguyễn Viết Hưng | | | | | | | | | | |  | | | | | | | | | | | | | | | |
| 1. **NHỮNG NGƯỜI THAM GIA THỰC HIỆN ĐỀ TÀI (NẾU CÓ)** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Họ và tên | | | | | | | Đơn vị công tác và lĩnh vực chuyên môn | | | | | | | | | Nội dung nghiên cứu cụ thể được giao | | | | | | | | Chữ ký | | |
| 2/ Lương Công Tâm  43.01.104.157 | | | | | | | Sinh viên khoa Công Nghệ Thông Tin | | | | | | | | |  | | | | | | | |  | | |
| 1. **ĐƠN VỊ PHỐI HỢP CHÍNH (NẾU CÓ)** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Tên đơn vị trong và ngoài nước | | | | | | | | | Nội dung phối hợp | | | | | | | | | | | | Họ và tên người đại diện | | | | | |
|  | | | | | | | | |  | | | | | | | | | | | |  | | | | | |
|  | | | | | | | | |  | | | | | | | | | | | |  | | | | | |
| 1. **TÌNH HÌNH NGHIÊN CỨU TRONG VÀ NGOÀI NƯỚC**    1. Tổng quan tình hình nghiên cứu thuộc lĩnh vực của đề tài | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Cùng với sự gia tăng tương tác người - máy (HCI), sự ra đời và phát triển của các thiết bị phần cứng hay phần mềm có khả năng giao tiếp với con người thông qua ngôn ngữ tự nhiên là một điều tất yếu về công nghệ hiện nay. Những thiết bị hay phần mềm có giao diện người dùng hỗ trợ giao tiếp qua ngôn ngữ tự nhiên như vậy được gọi chung là chatbot [1] [2]. Trong hội nghị Spring 2016, Facebook và Microsoft đã chính thức cung cấp các tài nguyên để xây dựng chatbots tích hợp vào nền tảng nhắn tin của họ là Messenger và Skype. Theo một công bố chính thức từ bài phỏng vấn của David Marcus, người đứng đầu mảng tin nhắn của Facebook, chỉ 6 tháng sau khi Facebook cung cấp tài nguyên để xây dưng chatbots trên nền tảng Messenger đã có đến khoảng 33000 nhà phát triển xây dựng khoảng 34000 chatbots dù chất lượng của chúng chưa thực sự tốt. Ngoài ra những nền tảng tin nhắn khác như Slack, Kik, Viber, … cũng đã hỗ trợ sự phát triển của chatbots.  Để có thể giao tiếp được với con người, chatbots cần phải trải qua quá trình xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP) để có thể đưa ra phản hồi tương ứng và có nghĩa cho con người. NLP xuất hiện từ rất sớm trong ngành khoa học máy tính. Từ những năm 1960, Weizenbaum đã xuất bản một nghiên cứu sáng tạo về tương tác ngôn ngữ tự nhiên với ELIZA, một chương trình máy tính được phát triển để bắt chước các phản ứng của nhà trị liệu tâm lý trong một buổi trị liệu tâm lý [3]. Ngoài ra, giải thưởng Loebner trong lĩnh vực trí tuệ nhân tạo cũng đóng góp không nhỏ thúc đẩy sự phát triển của NLP. Những tiến bộ vượt bậc những năm gần đây của Machine Learning hứa hẹn những cải tiến đáng kể trong khả năng diễn giải và dự đoán ngôn ngữ tự nhiên [4]. Các mô hình dự đoán trong NLP như Recurrent Neural Network (RNN) và Sequence-to-sequence ngày càng trở nên vượt trội hơn so với các mô hình luật dẫn (rule-based) truyền thống [5]. Tổng thể những tiến bộ về NLP trong thời gian gần đây đã nâng hiệu quả và chất lượng của chatbots lên một tầm cao mới.  Bên cạnh sự phát triển của NLP, các hệ trả lời câu hỏi dựa trên cơ sở tri thức (Knowledge-based Question Answering – KBQA) cũng đạt được nhiều bước tiến gần đây [6] [7] [8] [9]. Những bước tiến này góp phần hoàn thiện các chatbots giải đáp trả lời câu hỏi đặc biệt trong lĩnh vực giáo dục.  Hiện nay, chatbots sử dụng tiếng Việt chưa thực sự phổ biến, đặc biệt là chatbots trong lĩnh vực giáo dục. Hầu hết chatbots sử dụng tiếng Việt đều được thiết kế cho mục đích thương mại điện tử, marketing, chăm sóc khách hàng, … Chính vì sự khan hiếm của những nghiên cứu đi trước về một hệ trả lời câu hỏi bằng tiếng Việt nên đây cũng là một khó khăn lớn cần phải vượt qua của đề tài này. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| * 1. Danh mục các công trình liên quan (Họ và tên tác giả; Nhan đề bài báo, ấn phẩm; Các yếu tố về xuất bản)  1. Dale, R.: The return of the chatbots. Nat. Lang. Eng. 22(5), 811–817 (2016) 2. Følstad, A., Brandtzaeg, P.B.: Chatbots – the new world of HCI. ACM Interactions (2017, in Press) 3. Weizenbaum, J.: ELIZA – a computer program for the study of natural language communication between man and machine. Commun. ACM 9(1), 36–45 (1966) 4. Shah, H., Warwick, K., Vallverdú, J., Wu, D.: Can machines talk? Comparison of ELIZA with modern dialogue systems. Comput. Hum. Behav. 58, 278–295 (2016) 5. Vinyals, O., Le, Q.: A Neural Conversational Model. arXiv:1506.05869 (2015): 6. Mo Yu, Wenpeng Yin, Kazi Saidul Hasan, Cicero dos Santos, Bing Xiang, Bowen Zhou: Improved Neural Relation Detection for Knowledge Base Question Answering. arXiv:1704.06194v2 (2017): 7. Wanyun Cui, Yanghua Xiao, Haixun Wang, Yangqiu Song, Seung-won Hwang, Wei Wang: KBQA: Learning Question Answering over QA Corpora and Knowledge Bases. arXiv:1903.02419v1 (2019): 8. Lei, Kai & Deng, Yang & Zh, Bing & Shen, Ying. (2017). Open Domain Question Answering with Character-Level Deep Learning Models. 30-33. 10.1109/ISCID.2017.58. 9. Zhang, Wei Emma & Sheng, Quan & Lau, Jey & Abebe, Ermyas. (2017). Detecting Duplicate Posts in Programming QA Communities via Latent Semantics and Association Rules. 1221-1229. 10.1145/3038912.3052701. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1. **TÍNH CẤP THIẾT CỦA ĐỀ TÀI** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Trong thời đại cách mạng công nghiệp 4.0, định hướng STEM đang được triển khai như một hướng đi quan trọng trong giáo dục và đào tạo tại nhiều nước trên thế giới. Môn Tin học hội tụ đủ tất cả bốn yếu tố giáo dục STEM: Khoa học (S), công nghệ (T), kĩ nghệ (E) và toán học (M). Với tư cách là công nghệ nền tảng, tin học có vai trò trung tâm kết kết nối các môn học khác, qua đó đẩy mạnh giáo dục STEM.  Dự báo trong khoảng 10 năm tới, nhiều ngành nghề và việc làm hiện nay sẽ mất đi do sự thay thế nguồn lực lao động của máy tính, thay vào đó sẽ xuất hiện một số ngành nghề, việc làm mới đòi hỏi kiến thức, kĩ năng tin học chuyên sâu. Chương trình môn Tin học cần đáp ứng khả năng kết nối và lan toả sâu rộng của tin học sang tất cả các lĩnh vực khác nhau của đời sống, làm tốt định hướng nghề nghiệp.  Theo chương trình giáo dục phổ thông bộ môn tin học, ban hành ngày 26 tháng 12 năm 2018, phương pháp và định hướng giảng dạy có sự thay đổi lớn:   * **Đối với học sinh tiểu học và trung học cơ sở:** các em sẽ được tiếp cận và làm quen với máy tính và Internet, được rèn luyện tư duy Logic và tư duy giải quyết vấn đề với sự trợ giúp của máy tính thông qua việc xây dựng chương trình máy tính bằng các ngôn ngữ lập trình trực quan. * **Đối với học sinh trung học phổ thông:** các em sẽ được định hướng học tin học theo 2 chuyên đề “Tin học ứng dụng” và “Khoa học máy tính”. Nội dung chủ yếu xoay quanh việc sử dụng máy tính để giải quyết vấn đề thông qua việc lập trình và ứng dụng thuật toán. Việc học lập trình sẽ không bị bó buộc bởi một ngôn ngữ lập trình cụ thể mà sẽ do nhà trường, đơn vị giáo dục tại địa phương quyết định tùy theo điều kiện cụ thể.   Qua đó ta có thể thấy việc học lập trình đang được chú trọng và mở rộng chương trình ra nhiều cấp học, từ lập trình trực quan ở cấp tiểu học và trung học cơ sở tới lập trình ứng dụng ở cấp trung học phổ thông. Tuy nhiên, việc dạy và học lập trình ở Việt Nam vẫn còn tồn đọng một số khó khăn, hạn chế:   * Lực lượng giáo viên tin học còn ít, thiếu: Số lượng giáo viên hạn chế khiến cho các giáo viên hiện tại phải dạy hết công suất trong một tuần dẫn đến năng suất giảng dạy giảm, hay các vùng sâu vùng xa không được tiếp cận với tin học vì không đủ giáo viên hay thiết bị dạy học. * Tài liệu học lập trình bằng tiếng Việt còn ít: Đối với sinh viên chuyên ngành CNTT phải tiếp cận và tìm hiểu tài liệu bằng tiếng Anh là điều đương nhiên, tuy nhiên vẫn còn số lượng lớn sinh viên gặp khó khăn với ngoại ngữ, hay các học sinh cấp 2, cấp 3 muốn học lập trình sẽ gặp phải một rào cản rất lớn. * Thiếu sự tư vấn, giải đáp thắc mắc: Do số lượng giáo viên tin học còn thiếu, và với độ rộng của kiến thức trong phương pháp giáo dục phổ thông mới, học sinh sẽ gặp một số khó khăn trong việc tìm sự trợ giúp từ giáo viên, và giáo viên cũng khó có thể giải đáp kịp thời những thắc mắc của học sinh nếu như không được đào tạo giảng dạy theo đúng định hướng mà học sinh chọn.   Mục tiêu chính của đề tài này là góp phần thay đổi thực trạng trên với một hệ thống chatbot giải đáp thắc mắc trên nền Web cho học sinh, sinh viên, và giáo viên. Cụ thể, hệ thống sẽ giúp giải đáp những thắc mắc như: Các vấn đề về ngôn ngữ lập trình (C/C++, Python, nên bắt đầu học lập trình với ngôn ngữ nào, sự khác nhau giữa các ngôn ngữ, …), cung cấp kiến thức về cấu trúc dữ liệu và giải thuật (Thuật toán sắp xếp, tìm kiếm hoạt động ra sao, cách để giải quyết vấn đề bằng cách chia nhỏ vấn đề đó ra thành các vấn đề nhỏ hơn, tìm kiếm các hình thức biểu diễn trực quan để hiểu rõ cách hoạt động của thuật toán, …), giúp học sinh, sinh viên hiểu rõ hơn về lập trình hướng đối tượng... Các kiến thức sẽ được đi kèm theo các ví dụ cụ thể để người sử dụng có được cái nhìn tổng quan hơn về vấn đề mà họ đang thắc mắc.  Nhóm nghiên cứu hy vọng đề tài sẽ mang đến một công cụ học tập cho học sinh, sinh viên cho môn lập trình, giảm tải bớt phần nào áp lực cho các giáo viên tin học trong việc giải đáp các thắc mắc của học sinh về các lĩnh vực ngoài chuyên môn. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1. **MỤC TIÊU ĐỀ TÀI** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| * Xây dựng mô hình xử lý ngôn ngữ tự nhiên, ngôn ngữ tiếng Việt. * Xây dựng hệ thống giải đáp các kiến thức về ngôn ngữ lập trình lập trình, cấu trúc dữ liệu và giải thuật. * Xây dựng hệ thống chatbot, công cụ hỗ trợ trực quan, đa nền tảng cho các học sinh, sinh viên trong việc học lập trình. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1. **CÁCH TIẾP CẬN, PHƯƠNG PHÁP NGHIÊN CỨU, PHẠM VI NGHIÊN CỨU** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| * Xử lý ngôn ngữ tự nhiên (tiếng Việt) bằng mạng neural nhân tạo (Word-embedding/ RNN). * Biểu diễn tri thức lập trình cơ bản bằng một hệ cơ sở tri thức (Knowledge-based System) * Kết hợp NLP và KBS để xây dựng chatbot trả lời câu hỏi về lập trình cơ bản trên nền web * Phạm vi đề tài:   + Kiến thức lập trình với 2 ngôn ngữ C/C++ và Python với một số giải thuật cơ bản.   + Chatbot chỉ hỗ trợ giao tiếp qua ngôn ngữ tiếng Việt. | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1. **NỘI DUNG NGHIÊN CỨU VÀ TIẾN ĐỘ THỰC HIỆN** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| STT | Các nội dung, công việc  thực hiện chủ yếu | | | | | | | | | | | Sản phẩm  phải đạt | | | | | | Thời gian  (bắt đầu – kết thúc) | | | | | Người thực hiện | | | |
| 1 | Khảo sát các giáo viên, học sinh cấp 3 về các vấn đề trong việc học lập trình | | | | | | | | | | | Bảng tổng hợp, thống kê kết quả khảo sát | | | | | | 10/2019 – 10/2019 | | | | | Nhóm sinh viên | | | |
| 2 | Thu thập dữ liệu các kiến thức về lập trình, đánh giá dữ liệu | | | | | | | | | | |  | | | | | | 10/2019 – 11/2019 | | | | | Nhóm sinh viên | | | |
| 3 | Xây dựng cơ sở biểu diễn tri thức | | | | | | | | | | |  | | | | | | 11/2019 - 12/2019 | | | | | Nhóm sinh viên | | | |
| 4 | Nghiên cứu các mô hình NLP đã có | | | | | | | | | | | Bảng tổng hợp các ưu/nhược điểm và các đặc điểm đặc biệt của các loại models đã có | | | | | | 01/2020 | | | | | Nhóm sinh viên | | | |
| 5 | Lựa chọn hoặc xây dựng mô hình NLP phù hợp | | | | | | | | | | | Xây dựng model thích hợp đối với bài toán | | | | | | 01/2020 - 01/2020 | | | | | Nhóm sinh viên | | | |
| 6 | Thu thập, chuẩn bị dữ liệu huấn luyện mô hình | | | | | | | | | | | Các tập dataset cho quá trình train, test model | | | | | | 02/2020 | | | | | Nhóm sinh viên | | | |
| 7 | Thực hiện huấn luyện mô hình | | | | | | | | | | | Model đã được train | | | | | | 02/2020 | | | | | Nhóm sinh viên | | | |
| 8 | Kiểm thử và điều chỉnh thuật toán | | | | | | | | | | |  | | | | | | 02/2020 | | | | | Nhóm sinh viên | | | |
| 9 | Xây dựng hệ thống Chatbot trên nền Web | | | | | | | | | | |  | | | | | | 02/2020 | | | | | Nhóm sinh viên | | | |
| 10 | Triển khai, thử nghiệm tại khoa CNTT và một số trường trung học phổ thông | | | | | | | | | | |  | | | | | | 03/2020 | | | | | Nhóm sinh viên | | | |
| 11 | Khảo sát khả năng ứng dụng | | | | | | | | | | | Bảng khảo sát, đánh giá khả năng ứng dụng của đề tài trong các lĩnh vực liên quan | | | | | | 03/2020 | | | | | Nhóm sinh viên | | | |
| 12 | Tổng kết báo cáo | | | | | | | | | | | Bản báo cáo tổng hợp quá trình và kết quả nghiên cứu | | | | | | 04/2019 | | | | | Nhóm sinh viên | | | |
| 1. **SẢN PHẨM VÀ ĐỊA CHỈ ỨNG DỤNG** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| * **Loại sản phẩm:** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Mẫu | | | | ☐, | | Vật liệu | | | | | | ☐, | Thiết bị máy móc | | | | | | ☐ | | , Dây chuyền công nghệ | | | | | ☐ |
| Giống cây trồng | | | | ☐, | | Giống gia súc | | | | | | ☐, | Quy trình công nghệ | | | | | | ☐, | | Phương pháp | | | | | ☐ |
| Tiêu chuẩn | | | | ☐, | | Qui phạm | | | | | | ☐, | Sơ đồ | | | | | | ☐, | | Báo cáo phân tích | | | | | ☐ |
| Tài liệu dự báo | | | | ☐, | | Đề án | | | | | | ☐, | Luận chứng kinh tế | | | | | | ☐, | | Chương trình máy tính | | | | | ☒ |
| Bản kiến nghị | | | | ☐, | | Sản phẩm khác: | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | |
| * **Tên sản phẩm, số lượng và yêu cầu khoa học đối với sản phẩm** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **STT** | | **Tên sản phẩm** | | | | | | | | | | | | **Số lượng** | | | | **Yêu cầu khoa học** | | | | | | | | |
| 1 | | Model và bộ dữ liệu | | | | | | | | | | | | 1 | | | |  | | | | | | | | |
| 2 | | Báo cáo tổng kết | | | | | | | | | | | | 1 | | | |  | | | | | | | | |
| * **Địa chỉ có thể ứng dụng** (tên địa phương, đơn vị ứng dụng): | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1. **KINH PHÍ THỰC HIỆN ĐỀ TÀI** | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Tổng kinh phí: | | | | | | | | | | | | | | 2.500.000 VNĐ  (hai triệu năm trăm ngàn đồng) | | | | | | | | | | | | |
|  | | | | | | | | | | | | | |  | | | | | | | | | | | | |
| Dự trù kinh phí theo các mục chi: | | | | | | | | | | | | | |  | | | | | | | | | | | | |
| Chi phí Server, thiết bị huấn luyện AI | | | | | | | | | | | | | | 1.600.000 VNĐ | | | | | | | | | | | | |
| Chi phí nghiên cứu | | | | | | | | | | | | | | 900.000 VNĐ | | | | | | | | | | | | |
|  | | | | | | | | | | | | | |  | | | | | | | | | | | | |
|  | | | | | | | | | | | | | |  | | | | | | | | | | | | |
| *TPHCM, ngày \_\_ tháng \_\_ năm \_\_\_\_\_* | | | | | | | | | | | | | | *TPHCM, ngày \_\_ tháng \_\_ năm \_\_\_\_\_* | | | | | | | | | | | | |
| **Trưởng Khoa/Bộ môn**  (ký tên, đóng dấu) | | | | | | | | | | | | | | **GV hướng dẫn**  (Họ và tên, ký) | | | | | | | **Sinh viên thực hiện đề tài**  (Họ và tên, ký) | | | | | |
| *TPHCM, ngày \_\_ tháng \_\_ năm \_\_\_\_\_*  **HIỆU TRƯỞNG** | | | | | | | | | | | | | | | | | | | | | | | | | | |