



XÂY DỰNG KHUNG ỨNG DỤNG AI CHATBOT TRONG LĨNH VỰC QUY CHẾ ĐÀO TẠO

Đoàn Thị Hồng Phước*, Lê Văn Tường Lân, Nguyễn Văn Trung

Khoa Công nghệ Thông tin, Trường Đại học Khoa học, Đại học Huế, 77 Nguyễn Huệ, Huế, Việt Nam

Tóm tắt. Trong vài thập niên gần đây, trợ lý ảo thông minh (AI chatbot) đã và đang được nghiên cứu và phát triển trong nhiều lĩnh vực như thương mại điện tử, chăm sóc sức khỏe, du lịch và giáo dục do những thuận lợi của chúng trong việc tương tác giữa người và máy bằng ngôn ngữ tự nhiên. Quy chế đào tạo đại học giúp cho sinh viên đưa ra kế hoạch học tập phù hợp. Tuy nhiên, hiện tại sinh viên chỉ tiếp cận các quy chế này thông qua sổ tay sinh viên, các cố vấn học tập hoặc chuyên viên phòng Đào tạo. Việc này làm mất rất nhiều thời gian của sinh viên, cán bộ giảng dạy cũng như chuyên viên các phòng ban trong nhà trường. Để giải quyết vấn đề này, chúng tôi đề xuất một khung ứng dụng AI chatbot dựa trên nền tảng Rasa và áp dụng thử nghiệm vào lĩnh vực quy chế đào tạo đại học. Kết quả chạy thử nghiệm chương trình cho thấy rằng độ chính xác của mô hình hiểu ngôn ngữ tự nhiên là 84,6% và độ chính xác theo đánh giá của các chuyên gia là 89,3%.

Từ khóa: trợ lý ảo, AI chatbot, quy chế đào tạo đại học, Rasa

Developing AI chatbot framework in training regulation domain

Doan Thi Hong Phuoc*, Le Van Tuong Lan, Nguyen Van Trung

Faculty of Information Technology, University of Sciences, Hue University, 77 Nguyen Hue St., Hue, Vietnam

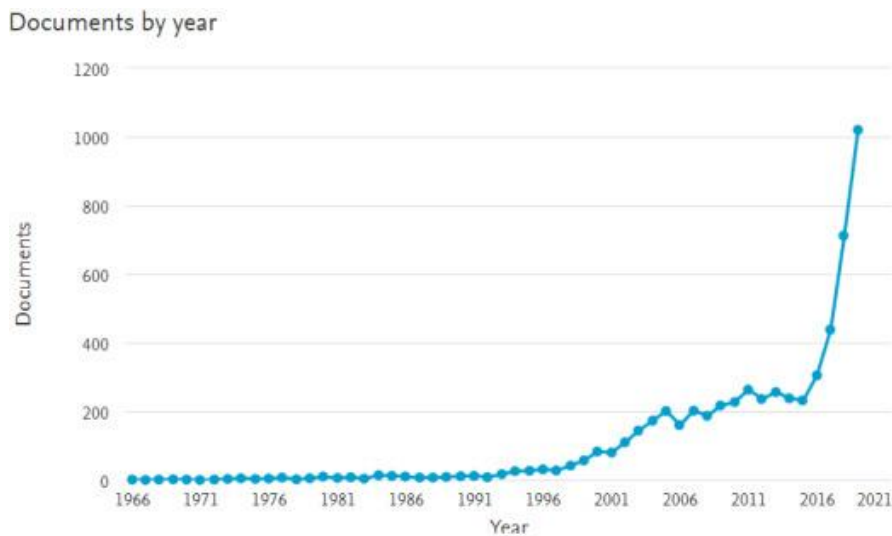
Abstract. In recent decades, AI chatbots (aka artificial intelligence virtual assistants) have been investigated and developed in numerous domains, such as e-commerce, health care, tourism, and education, because of their benefits in the interaction between humans and machines via natural language text/speech. Universities' training regulation is necessary for students to plan their studies. However, students can only access these rules via the training handbook, their training counsellors, or the staff of the Office for undergraduate education. This process is time-consuming for students, lecturers, and office staff. To deal with this problem, we proposed an AI chatbot framework based on the Rasa platform and applied it to the university's training regulation domain. The results show that the F-measure of the natural language understanding model is 84.6%, and the experts' accuracy evaluation is 89.3%.

* Liên hệ: dthphuoc@husc.edu.vn

Keywords: AI chatbot, university training regulations, Rasa

1 Mở đầu

Chatbot là một chương trình máy tính, phản hồi giống như một thực thể thông minh khi trò chuyện thông qua văn bản hoặc giọng nói và hiểu một hoặc nhiều ngôn ngữ của con người bằng xử lý ngôn ngữ tự nhiên [1]. Chatbot hỗ trợ các tổ chức kinh tế, xã hội bằng cách thay thế các nhân viên trả lời câu hỏi người dùng mọi lúc mọi nơi, giúp giảm thời gian và chi phí cho người dùng. Theo các kết quả nghiên cứu trong Scopus (thống kê theo các từ khóa “chatbot”, “tác tử trò chuyện” (conversation agent) và “giao diện trò chuyện” (conversation interface)) (Hình 1), có thể thấy rằng từ năm 2016 trở về sau, các ứng dụng Chatbot đã có sự gia tăng nhanh chóng với các loại hình hệ thống Chatbot khác nhau được sử dụng cho các lĩnh vực khác nhau như chăm sóc sức khỏe, du lịch, thương mại điện tử và giáo dục [2].



Hình 1. Biểu đồ mô tả số lượng tài liệu nghiên cứu liên quan đến lĩnh vực Chatbot từ năm 1966 đến 2021 [2]

Hiện nay, các cơ sở giáo dục cũng như các trường đại học (ĐH) trên giới đã triển khai các ứng dụng trợ lý ảo thông minh (AI chatbot) hỗ trợ cho các hoạt động trong nhà trường. Trợ lý ảo Deakin Genie của Trường ĐH Deakin (Deakin's Genie assistant) hỗ trợ chuyên viên nhà trường trả lời các câu hỏi của sinh viên liên quan đến lịch học, bài tập, chính sách học bổng, cho vay và các nguồn tài nguyên trong trường [3]. Amazon Alexa, phát triển từ mã nguồn mở QnABot của Amazon, là một trợ lý ảo được sử dụng tại ĐH St. Louis. Alexa có thể thay thế các nhà quản lý nội trú trả lời các câu hỏi các sự kiện và tổ chức của sinh viên cùng với nhiều khía cạnh khác của

cuộc sống sinh viên [4]. ĐH Georgia State đã xây dựng trợ lý ảo thông minh Pounce nhằm giúp sinh viên bằng cách gửi lời nhắc về thời gian và các thông tin liên quan đến các công việc đăng ký học (enrolment tasks), thu thập dữ liệu khảo sát và trả lời ngay lập tức nhiều câu hỏi của sinh viên bất kể thời gian nào [5]. ĐH Khoa học ứng dụng Amsterdam (Amsterdam University of Applied Sciences – AUAS) đã tích hợp chatbot Robin trên trang web của Trường nhằm mục đích trả các câu hỏi thường gặp của sinh viên [6].

Tại Việt Nam, trong hơn một thập niên qua, AI chatbot đã được nghiên cứu, xây dựng và áp dụng tại một số trường đại học. Nhóm nghiên cứu tại Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hồ Chí Minh, đã xây dựng chatbot FIT-Ebot để hỗ trợ cho việc quản lý và học tập [7]. Nhóm tác giả của Trường Đại học Cần Thơ và Trường Đại học Nguyễn Tất Thành đề xuất xây dựng AI chatbot trên máy tính Raspberry Pi có thể trả lời tự động cho sinh viên ngành Công nghệ thông tin các câu hỏi liên quan đến môi trường học tập và phương pháp học tập bậc đại học, kỹ năng nghề nghiệp và xu hướng công nghệ [8]. Nhóm tác giả của Học viện Bưu chính viễn thông đã đưa ra trợ lý ảo thông minh trợ giúp sinh viên PTIT chatbot, có khả năng giải đáp các thắc mắc của sinh viên mới trên các lĩnh vực về nhà trường và đời sống [9]. Gần đây, chatbot NEU hỗ trợ cho việc tuyển sinh đã được đưa vào hoạt động trên trang “Tư vấn tuyển sinh đại học chính quy” của Trường Đại học Kinh tế Quốc dân [10].

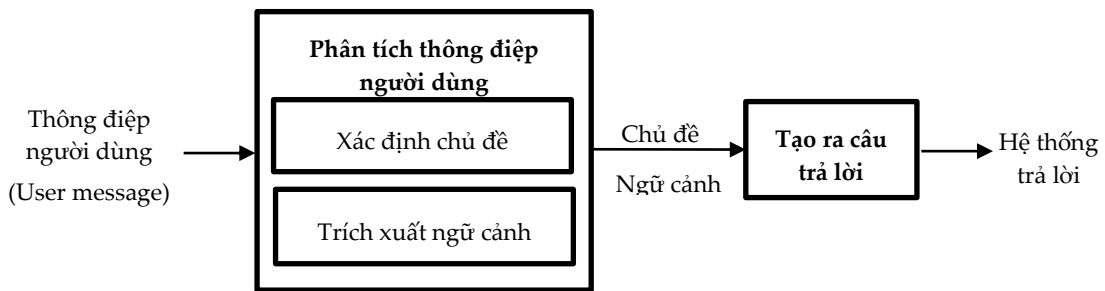
Trên thực tế, các thông tin về quy chế đào tạo đại học rất cần thiết để sinh viên có thể đưa ra kế hoạch học tập phù hợp. Tuy nhiên, hiện tại sinh viên chỉ tiếp cận các quy chế này thông qua sổ tay sinh viên, cổ vấn học tập hoặc chuyên viên phòng đào tạo. Ngoài ra, các câu hỏi về các quy chế này thường được lặp lại từ năm này qua năm khác. Việc này làm mất rất nhiều thời gian của sinh viên, cán bộ giảng dạy cũng như chuyên viên các phòng ban trong nhà trường. Tại Việt Nam có rất ít AI chatbot thiết kế cho lĩnh vực đào tạo. Gần đây, nhóm tác giả trong nghiên cứu [7] đã xây dựng trợ lý ảo FIT-Ebot cho lĩnh vực đào tạo dựa trên nền tảng Dialogflow của Google và là nền tảng đóng, rất khó tích hợp và tùy chỉnh. Vì vậy, trong bài báo này, chúng tôi xây dựng một khung ứng dụng AI chatbot và áp dụng vào lĩnh vực quy chế đào tạo đại học trên nền tảng Rasa Framework với khả năng tích hợp và tùy chỉnh một cách dễ dàng.

Các phần tiếp theo của bài báo được tổ chức như sau: Phần 2 trình bày cách tiếp cận xây dựng khung ứng dụng AI chatbot. Phần 3 trình bày quy trình cài đặt thử nghiệm AI chatbot trong lĩnh vực quy chế đào tạo của Trường Đại học Khoa học, Đại học Huế. Phần 4 trình bày kết luận và hướng phát triển.

2 Cách tiếp cận xây dựng khung ứng dụng AI chatbot

2.1 Cấu trúc chung của AI chatbot

Hiện nay, có nhiều loại AI chatbot khác nhau và được phân loại dựa trên một số tiêu chí như cách thức tương tác (interaction mode), miền tri thức (knowledge domain), loại ứng dụng (chatbot application) và cách tiếp cận để xây dựng (design approach) [11]. Tuy nhiên, nhìn chung, cấu trúc của một hệ trợ lý ảo gồm hai thành phần chính: (1) Phân tích thông điệp người dùng; (2) Tạo ra câu trả lời tương ứng với thông điệp đầu vào:



Hình 2. Cấu trúc chung của một trợ lý ảo [7]

(1) Phân tích thông điệp người dùng (User message analysis): Giúp hệ thống hiểu rõ những gì mà người sử dụng đang yêu cầu, gồm hai chức năng:

Xác định chủ đề của người dùng (Identifying user intent): Mục đích của chức năng này là xác định chủ đề (ý định) của người dùng thông qua thông điệp mà họ đưa vào. Một số cách tiếp cận chính cho việc xác định chủ đề bao gồm đối sánh mẫu, dựa trên luật và học máy. Đối với cách tiếp cận học máy, bài toán xác định chủ đề của câu văn bản cho trước được xem như là bài toán phân loại văn bản, trong đó mỗi văn bản được phân vào lớp chủ đề tương ứng. Một số phương pháp học máy được sử dụng như SVM, Naive Bayes, logistic regression và các phương pháp học sâu [12].

Trích xuất ngữ cảnh người dùng (Extracting user context): Chức năng này có nhiệm vụ trích xuất các thông tin ngữ cảnh thông qua các thực thể (entities) của thông điệp người dùng, chẳng hạn như địa điểm, thời gian và thông tin người dùng. Những thông tin này giúp hệ thống đưa ra câu trả lời phù hợp với tình huống của người dùng.

Ví dụ, với câu “Giảng viên giảng dạy học phần Trí tuệ nhân tạo?”, hệ thống xác định được chủ đề người dùng muốn hỏi liên quan đến thông tin học phần. Để có thể đưa ra phản hồi phù hợp, hệ thống cần hiểu hơn về thông điệp này. Hệ thống đòi hỏi thêm về thông tin ngữ cảnh như:

môn học (“Trí tuệ nhân tạo”) và *giảng viên*. Trong trường hợp không trích xuất được thêm thông tin ngữ cảnh, hệ thống yêu cầu người dùng nhập thêm vào.

(2) **Tạo ra câu trả lời (Response generation):** Thành phần này tạo ra câu trả lời phù hợp dựa vào chủ đề và ngữ cảnh thông điệp của người dùng. Có ba cách tiếp cận chính:

Dựa vào mẫu (Pattern-based approach): Phương pháp này sẽ so khớp thông điệp đầu vào với mẫu câu hỏi – đáp để đưa ra câu trả lời tương ứng

Mô hình truy hồi (Retrieval-based model approach): Dựa vào chủ đề và ngữ cảnh của thông điệp đầu vào, hệ thống sẽ đưa ra câu trả lời thích hợp bằng cách truy vấn vào cơ sở dữ liệu/cơ sở tri thức.

Mô hình sinh (Generative-based model approach): Đưa ra câu trả lời gần giống cách của con người nhất. Cụ thể, hệ thống sẽ đưa ra câu trả lời dựa vào các thông điệp đầu vào hiện tại và trước đó. Tuy nhiên, trên thực tế, cách tiếp cận này gặp khó khăn trong việc xây dựng và huấn luyện mô hình vì dữ liệu huấn luyện cần phải rất lớn để có thể đưa ra câu trả lời chính xác.

2.2 Cách tiếp cận xây dựng AI chatbot

Việc xây dựng AI chatbot dựa trên framework sẵn có giúp tiết kiệm thời gian và kinh phí rất nhiều. Hiện tại, Rasa framework [13] có nhiều ưu điểm hơn một số framework hỗ trợ cho việc xây dựng AI chatbot như ChatterBot, Dialogflow, Wit.ai và FPT.AI. Thật vậy, Rasa là một khung chatbot mã nguồn mở cho phép tích hợp và tùy chỉnh các thành phần một cách dễ dàng. Ngoài ra, Rasa hỗ trợ kết nối chương trình với các ứng dụng tin nhắn khác và có thể triển khai trên nhiều môi trường khác nhau.

Rasa có hai thành phần chính là Rasa Core và Rasa NLU. Rasa NLU có chức năng phân tích chủ đề/ý định và trích xuất ngữ cảnh của thông điệp thông qua các thực thể. Rasa Core được sử dụng để quản lý luồng hội thoại (Dialogue Manager), phát biểu (utterance) và hành động (action). Nói cách khác, Rasa NLU là thành phần hiểu/điễn giải ý định của thông điệp đầu vào và Rasa Core đưa ra phản hồi (phát biểu/hành động) tương ứng tại mỗi bước của hội thoại và dự đoán trạng thái tiếp theo dựa vào một mô hình xác suất và tùy thuộc vào chính sách lựa chọn. Rasa Core cung cấp hai loại chính sách: chính sách học máy (Machine learning policy) và chính sách dựa trên luật (Rule-based policy):

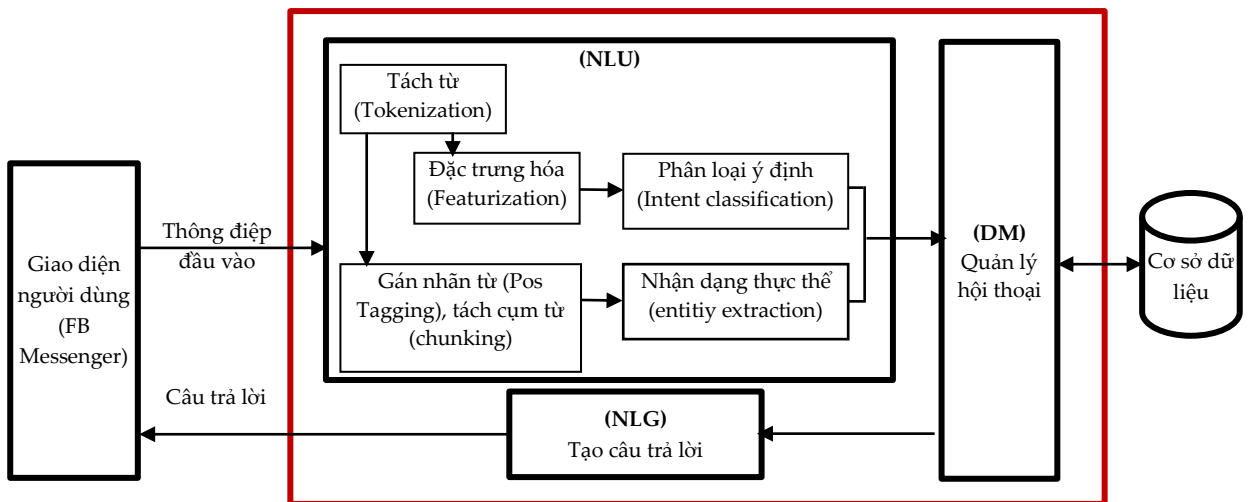
- **Chính sách học máy:**

- *Chính sách TED (Transformer Embedding Dialogue policy):* Chính sách này kết hợp các đặc trưng đầu vào người sử dụng (input user), hành động trước đó của hệ thống (system

actions) và dữ liệu trong vùng nhớ của chatbot (slots) tại mỗi bước thành vec to đầu vào cho việc mã hóa bộ chuyển đổi hội thoại (dialogue transformer encoder).

- *Chính sách ghi nhớ (Memoization policy)*: Ghi nhớ các câu chuyện (stories) từ tập dữ liệu huấn luyện. Kiểm tra câu chuyện được so khớp (matching story) của cuộc hội thoại hiện tại và dự đoán hành động tiếp theo từ câu chuyện so khớp với độ tin cậy trong đoạn $[0,1]$. Số lượt hội thoại được xác định trong tham số *max-history*.
 - *Chính sách ghi nhớ tăng cường (Augmented Memoization policy)*: Ghi nhớ các mẫu từ câu chuyện so khớp tùy theo số lượt *max-history*. Tương tự chính sách *memoization policy*, chính sách này có cơ chế quên.
- **Chính sách dựa trên luật:**
 - Là chính sách giải quyết các phần của hội thoại theo cách hành xử cố định sẵn và đưa ra các dự đoán bằng việc sử dụng các quy tắc/luật đã được định nghĩa trong tập training.

Dựa trên nền tảng Rasa, chúng tôi xây dựng khung AI chatbot, gồm các thành phần chính trình bày trên Hình 3.



Hình 3. Khung AI chatbot

Khung này bao gồm ba thành phần chính: (1) Facebook Messenger dùng để giao tiếp với người dùng (nhận thông điệp đầu vào và hiển thị phản hồi của hệ thống); (2) Nền tảng RASA Framework dùng để xác định chủ đề, trích xuất thông tin từ thông điệp đầu vào của người dùng

và đưa ra câu trả lời tương ứng tại mỗi bước của hội thoại và dự đoán trạng thái tiếp theo (quản lý luồng hội thoại); (3) Cơ sở dữ liệu dùng để tham khảo và tạo câu trả lời.

Thông điệp của người dùng sau khi tiếp nhận sẽ được chuyển đến thành phần diễn giải (Rasa NLU) để xác định chủ đề và trích xuất các thông tin cần thiết. Cụ thể, thông điệp đầu vào sẽ được xử lý sơ bộ thông qua các bước xử lý ngôn ngữ tự nhiên như tách từ, chuyển câu văn bản thành vec to đặc trưng (đặc trưng hóa) để đưa vào bộ phân lớp nhận dạng chủ đề và gán nhãn từ và trích xuất các cụm từ để nhận diện các thực thể. Dựa vào các thông tin về chủ đề và ngữ cảnh được trích xuất, thành phần quản lý luồng hội thoại (Rasa Core) đưa ra câu trả lời (domain.yml) phù hợp với thông điệp đầu vào và đưa ra hành động tiếp theo tùy vào cấu hình chính sách lựa chọn (config.yml): dựa trên mô hình đã học trong tập training (nlu.md) hay đối sánh trong tập luật (rule.yml) hoặc cả hai.

3 Cài đặt thử nghiệm khung AI chatbot

Sau đây là phần mô tả cài đặt khung AI chatbot, áp dụng cho lĩnh vực quy chế đào tạo của Trường Đại học Khoa học, Đại học Huế.

3.1 Thiết kế chương trình

– *Giao tiếp với người dùng*: Phần giao tiếp với người dùng được cài đặt bằng cách sử dụng API tin nhắn của Facebook [14, 15]. Chi tiết của việc áp dụng này là như sau: Chúng tôi đã thiết lập một trang Facebook và cài đặt ứng dụng để nhận thông tin đăng nhập từ nhà phát triển Facebook và đưa thông tin này vào file credentials.yml. Sau đó, chèn URL gọi lại “https://<HOST>/webhooks/facebook/webhook”, xác nhận định danh (token), mã bí mật ứng dụng (App Secret) và định danh trang truy cập (page access token) vào file credential.yml.

– *Xây dựng tập dữ liệu huấn luyện (training data)*: Tập dữ liệu được sử dụng cho quá trình học mô hình xác định chủ đề và thực thể của thông điệp đầu vào. Việc thu thập dữ liệu huấn luyện được tiến hành như sau:

Chúng tôi đã thiết kế form¹ thu thập các câu hỏi thường gặp của sinh viên liên quan đến quy chế đào tạo đại học. Kết quả thu được hơn 500 mẫu câu hỏi liên quan đến năm chủ đề (intents) chính: điểm tích lũy, học phần, học bổng, thông tin tốt nghiệp và những vấn đề liên quan khác.

¹ <https://docs.google.com/forms/d/e/1FAIpQLSeFSvo0YDAQwc6SfemIEZuP4AG16i07G6bw6J6v-0R76LGJQw/viewform>

Từ tập dữ liệu thu được, chúng tôi tiến hành phân loại (gán nhãn) các câu hỏi vào hơn 50 chủ đề khác nhau liên quan đến quy chế đào tạo bằng cách chia nhỏ (làm mịn) năm chủ đề chính (Bảng 1).

Ngoài ra, để có thể học được ngữ cảnh, cần phải xác định các thực thể (entities) trong các câu hỏi (nếu có) của tập dữ liệu thu thập được. Cụ thể, các thực thể được trích xuất bằng cách sử dụng công cụ NER có sẵn [16]. Tuy nhiên, các công cụ có sẵn chỉ có thể trích xuất được các thực thể thông dụng cho nhiều lĩnh vực như địa điểm, tổ chức, thời gian, tên riêng và số lượng. Vì vậy, trong bài báo này, chúng tôi sử dụng công cụ có sẵn kết hợp với xác định bằng tay các thực thể trong lĩnh vực quy chế đào tạo đại học, như “Toán học rời rạc” là một thực thể “môn học”/”học phần”.

Bảng 1. Dạng câu hỏi và chủ đề tương ứng

Câu hỏi	Chủ đề (intent)
Cách thức đăng ký học phần bị trễ hạn	Đăng ký học phần trễ hạn
Tôi muốn hủy học phần đã đăng ký	Hủy học phần
Điều kiện để nhận học bổng là gì	Điều kiện học bổng
Học phí của trường này khoảng bao nhiêu?	Học phí
....	

Ví dụ, câu văn bản “*Học phần Toán rời rạc có bao nhiêu tín chỉ?*” có các thực thể “*Toán rời rạc*” và “*tín chỉ*” và được gán nhãn chủ đề “*Thông tin học phần*”.

```

- story: sotinchicuamoimonhoc
  steps:
  - intent: sotinchicuamoimonhoc
    entities:
    - subjects:
    - action: action_sotinchicuamonhoc
- story: hocbongdauvao
  steps:
  - intent: hocbongdauvao
  - action: utter_hocbongdauvao
- story: hocbongkhuyenkhich
  steps:
  - intent: hocbongkhuyenkhich
  - action: utter_hocbongkhuyenkhich

```

Hình 4. Một ví dụ về kịch bản liên quan đến thông tin số tín chỉ của môn học và thông tin học bổng

– **Tạo câu trả lời:** Trong bài báo này, chúng tôi sử dụng cách tiếp cận dựa vào mẫu (Pattern-based approach) và truy hồi thông tin (Retrieval-based approach). Cụ thể, trong chương trình có hai loại câu trả lời: (1) Định nghĩa các câu trả lời cố định là các phát biểu (utterance) dựa vào văn bản quy chế đào tạo đại học của Trường Đại học Khoa học², Đại học Huế, bao gồm 5 chương với 29 điều; (2) Xây dựng cơ sở dữ liệu chứa các thông tin liên quan đến quy chế đào tạo. Dựa vào chủ đề và ngữ cảnh của thông điệp đầu vào, hệ thống sẽ đưa ra câu trả lời thích hợp bằng cách truy vấn vào cơ sở dữ liệu.

– **Thiết lập cấu hình cho luồng (pipeline) công việc trong file config.yml:** Bao gồm các thành phần tạo nên NLU pipeline và các cơ chế phản hồi (Rasa Core) (Hình 5 và Hình 6).

- Cấu hình NLU pipeline

Bên cạnh cho phép người sử dụng tích hợp các công cụ tiền xử lý từ bên ngoài, Rasa cung cấp một số mô hình ngôn ngữ như MitieNLP, SpacyNLP và HFTransformersNLP cho việc tiền xử lý như tách từ, vec tơ hóa đặc trưng (biểu diễn ngôn ngữ). Trong bài báo này, chúng tôi sử dụng mô hình ngôn ngữ SpacyNLP và cấu hình thông số để biểu diễn văn bản tiếng việt [17].

```
language: vi
pipeline:
  - name: SpacyNLP
    model: "vi_spacy_model"
  - name: SpacyTokenizer
  - name: SpacyFeaturizer
  - name: RegexFeaturizer
  - name: LexicalSyntacticFeaturizer
  - name: CountVectorsFeaturizer
  - name: CountVectorsFeaturizer
    analyzer: char_wb
    min_ngram: 1
    max_ngram: 4
  - name: DIETClassifier
    epochs: 100
    constrain_similarities: true
  - name: EntitySynonymMapper
  - name: FallbackClassifier
    threshold: 0.5
    ambiguity_threshold: 0.1
```

Hình 5. NLU pipeline

² <https://it.husc.edu.vn/vanban/daihoc/?detail=227>

Hiện tại Rasa hỗ trợ một số bộ phân lớp chủ đề như KeywordIntentClassifier, MitieIntentClassifier, SklearnIntentClassifier và DIETClassifier. Trong đó, KeywordIntentClassifier là phương pháp đối sánh từ khoá dành cho dữ liệu nhỏ; bộ phân lớp MitieIntentClassifier sử dụng phương pháp phân lớp SVM đa lớp tuyến tính và SklearnIntentClassifier sử dụng phương pháp SVM tuyến tính để phân lớp chủ đề; DIETClassifier (Dual Intent and Entity Transformer) là bộ phân lớp với kiến trúc đa nhiệm vụ gồm phân lớp chủ đề và nhận diện thực thể có trong câu văn bản. DIETClassifier là phương pháp hiện đại nhất, cải tiến từ các phương pháp trước đây và cho ra kết quả tốt nhất trong các phương pháp xác định chủ đề hiện nay [18]. Ngoài ra, Rasa cung cấp bộ phân lớp FallbackClassifier để phân loại câu văn bản vào chủ đề *nlu_fallback* trong trường hợp việc phân lớp nhập nhằng, nghĩa là dựa vào ngưỡng độ tin cậy *fallback_threshold* để xem xét chủ đề của câu văn bản thuộc vào tập các chủ đề đã có hay thuộc vào chủ đề *nlu_fallback*. Trong nghiên cứu này, chúng tôi kết hợp các bộ phân lớp DIETClassifier và FallbackClassifier.

- *Cấu hình Rasa Core*: Cấu hình các chính sách để quyết định hành động cần thực hiện tại mỗi bước trong cuộc trò chuyện. Đa chính sách được lựa chọn để thiết kế trong chương trình này.

```

policies:
  - name: MemoizationPolicy
  - name: TEDPolicy
    max_history: 5
    epochs: 100
    constrain_similarities: true
  - name: RulePolicy
    core_fallback_threshold: 0.3
    core_fallback_action_name: "action_default_fallback"
    enable_fallback_prediction: True

```

Hình 6. Rasa Core pipeline

3.2 Kết quả thực nghiệm

Trong bài báo này, độ đo được sử dụng để đánh giá mô hình Rasa NLU và đánh giá của chuyên gia về các câu trả lời tương ứng với câu hỏi người sử dụng đưa vào khi triển khai chương trình là *Precision*, *Recall*, *F-measure (F1)* và *Accuracy* tương ứng ở các công thức (1), (2), (3) và (4).

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

trong đó *TP* (*True Positive*) là tổng số trường hợp dự báo đúng với lớp dương; *TN* (*True Negative*) là tổng số trường hợp dự báo đúng với lớp âm; *FP* (*False Positive*) là tổng số trường hợp các quan sát có nhãn thuộc lớp âm được dự báo lớp dương; *FN* (*False Negative*) là tổng số trường hợp các quan sát có nhãn thuộc lớp dương được dự báo lớp âm.

– *Đánh giá mô hình Rasa NLU*

Sử dụng phương pháp k-fold cross validation với $k = 5$ bằng câu lệnh như sau: `rasa test nlu --nlu data/nlu --cross-validation --folds 5`. Kết quả đánh giá trên tập test đối với bộ phân lớp các chủ đề (intent classification) như sau: Accuracy = 84,6%; F1-score = 82,6%; Precision: 83,6% với ma trận nhầm lẫn trình bày trên Hình 7.

– *Đánh giá của chuyên gia*

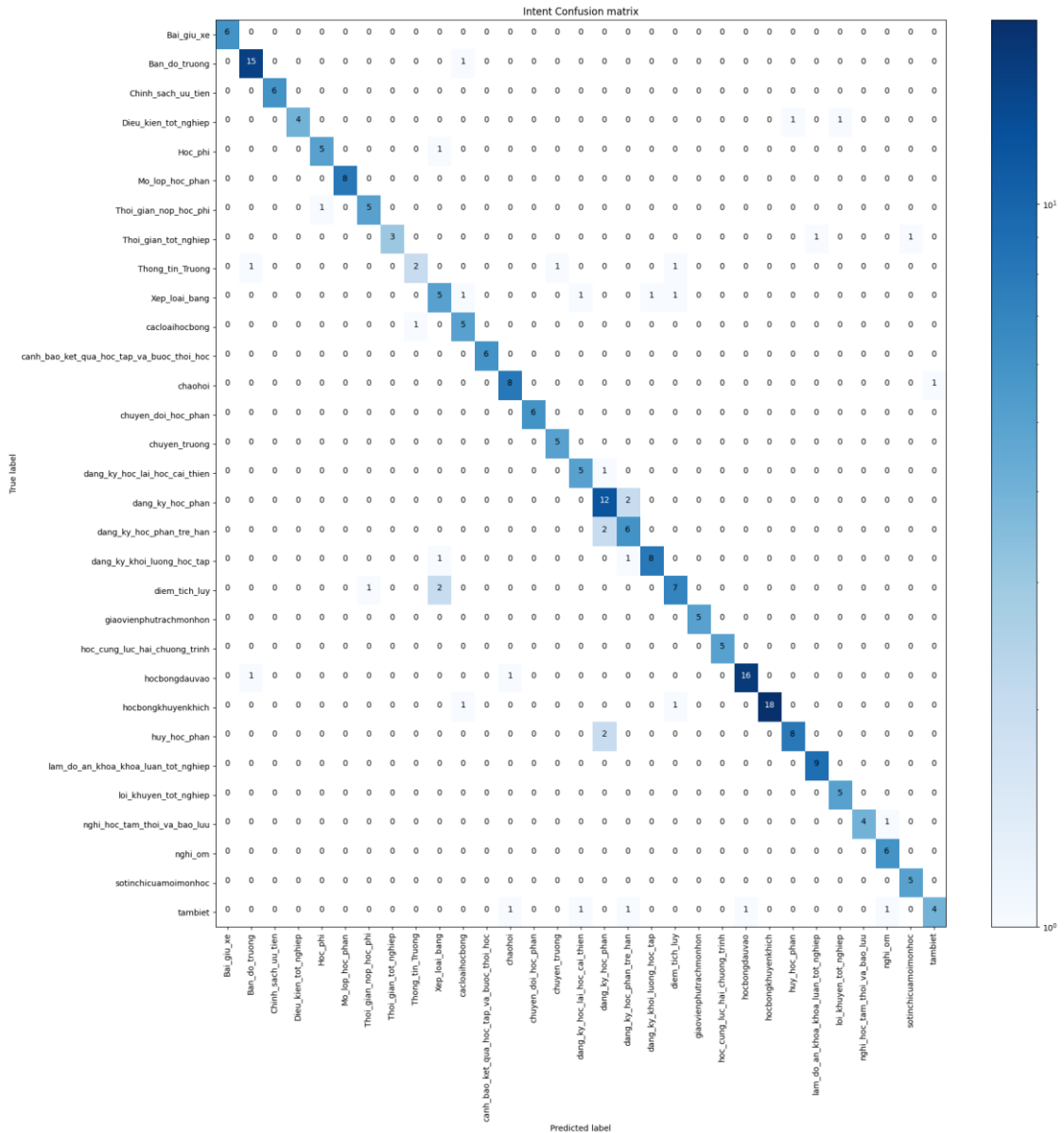
Chúng tôi đã tạo trang Facebook “Hỗ trợ đào tạo – ĐHKH” và kết nối với chương trình thông qua trình tin nhắn Facebook để chạy thử nghiệm. Chúng tôi mời một số chuyên gia (giảng viên, sinh viên, nhân viên phòng đào tạo) đặt câu hỏi thông qua cửa sổ chat FaceBook. Kết quả có 89,3% câu trả lời đúng với câu hỏi yêu cầu của chuyên gia. Nội dung các câu hỏi liên quan đến lĩnh vực quy chế đào tạo đại học như cách đăng ký học phần, điều kiện được học bổng, thông tin các học phần,...như được mô tả ở Hình 8. Các trường hợp đưa ra câu trả lời sai do người sử dụng đưa vào câu hỏi có chữ viết tắt, sai chính tả hoặc trong câu văn có từ đa nghĩa. Ngoài ra, đối với những câu hỏi không liên quan đến lĩnh vực đào tạo hoặc không có trong cơ sở dữ liệu thì bot sẽ đưa ra câu trả lời “Không hiểu câu hỏi của bạn”.

4 Kết luận và hướng phát triển

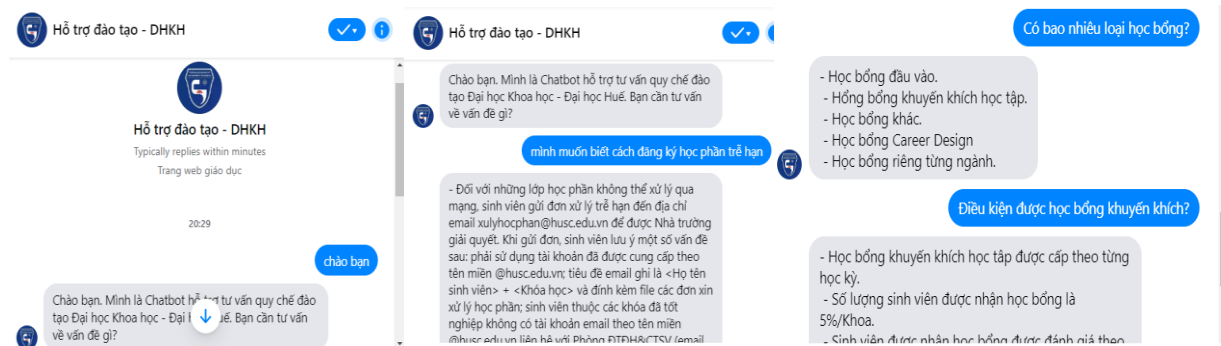
Bài báo trình bày khung AI chatbot dựa trên Rasa và minh họa cài đặt cho một lĩnh vực cụ thể, đó là lĩnh vực quy chế đào tạo đại học tại Trường Đại học Khoa học, Đại học Huế. Bản cài đặt cùng các thông số được thiết lập cho ra kết quả khá tốt. Mô hình hiểu ngôn ngữ tự nhiên có độ chính xác 84,6% và độ chính xác theo đánh giá của các chuyên gia là 89,3%.

Trên thực tế, chương trình thử nghiệm cho ra kết quả tốt khi câu văn bản đầu vào là một câu đầy đủ về cú pháp và không có ký tự viết tắt. Hướng phát triển của chương trình cần là phải xây dựng phần tiền xử lý tiếng Việt. Ngoài ra, chương trình cần xây dựng một cơ sở tri thức để

có thể suy diễn và đưa ra gợi ý các câu trả lời khi thông điệp đầu vào không có trong tập dữ liệu đã định nghĩa sẵn.



Hình 7. Ma trận nhầm lẫn



Hình 8. Một số hình ảnh minh họa cuộc hội thoại giữa người sử dụng và bot “Hỗ trợ đào tạo – ĐHKH”

Tài liệu tham khảo

1. A. Khanna, B. Pandey, K. Vashishta, K. Kalia, B. Pradeepkumar, and T. Das, “A Study of Today’s A.I. through Chatbots and Rediscovery of Machine Intelligence,” *Int. J. U- E-Serv. Sci. Technol.*, vol. 8, no. 7, pp. 277–284, Jul. 2015, doi: 10.14257/ijunesst.2015.8.7.28.
2. E. Adamopoulou and L. Moussiades, “Chatbots: History, technology, and applications,” *Mach. Learn. Appl.*, vol. 2, 2020, doi: 10.1016/j.mlwa.2020.100006.
3. Matt Johnston, “Deakin’s Genie assistant tackles 12,000 conversations a day,” 2019. [Online]. Available: <https://www.itnews.com.au/news/deakins-genie-assistant-tackles-12000-conversations-a-day-530688>
4. “Building a multi-channel Q&A chatbot at Saint Louis University using the open source QnABot,” 2019. <https://aws.amazon.com/blogs/publicsector/building-a-multi-channel-qa-chatbot-at-saint-louis-university-using-the-open-source-qnabot/>
5. “Georgia State University supports every student with personalized text messaging,” 2016. <https://mainstay.com/case-study/how-georgia-state-university-supports-every-student-with-personalized-text-messaging/>
6. JOOST, “Amsterdam University Launches Chatbot.” <https://cy2.nl/amsterdam-university-launches-chatbot/>
7. H. Hien, C. Pham-Nguyen, L. Nam, H. Nhung, and T. Le Dinh, “Intelligent Assistants in Higher-Education Environments: The FIT-EBot, a Chatbot for Administrative and Learning Support,” in *Proceedings of the Ninth International Symposium on Information and Communication Technology*, Danang City, Vietnam, 2018, pp. 69–76. doi: 10.1145/3287921.3287937.
8. Nghị Đ. T. and Tùng H., “Chatbot cho sinh viên công nghệ thông tin,” in *Kỷ yếu hội nghị khoa học công nghệ quốc gia lần thứ XII nghiên cứu cơ bản và ứng dụng công nghệ thông tin*, Hanoi, Vietnam, Oct. 2019. doi: 10.15625/vap.2019.00012.

9. "https://dantri.com.vn/nhan-tai-dat-viet/san-pham-tro-ly-ao-thong-minh-tro-giup-sinh-vien-ptit-chatbot-201908201268409412."
10. T. T. Nguyen, A. D. Le, H. T. Hoang, and T. Nguyen, "NEU chatbot: Chatbot for admission of National Economics University," *Comput. Educ. Artif. Intell.*, vol. 2, p. 100036, Jan. 2021, doi: 10.1016/j.caeai.2021.100036.
11. S. Hussain, O. Sianaki, and N. Ababneh, "A Survey on Conversational Agents/Chatbots Classification and Design Techniques," 2019, pp. 946–956. doi: 10.1007/978-3-030-15035-8_93.
12. Đoàn Thị Hồng Phước, Nguyễn Văn Trung, and Lê Văn Tường Lâm, "Một số phương pháp xác định chủ đề của câu văn bản trong hệ tư vấn," *Tạp Chí KHCN Trường Đại Học Khoa Học Huế*, vol. 18, no. 1, p. 47, Jun. 2021.
13. "Introduction to Rasa Open Source." [Online]. Available: <https://rasa.com/docs/rasa/>
14. "Connecting to Facebook messenger." <https://rasa.com/docs/rasa/connectors/facebook-messenger/>
15. "Facebook for Developer." <https://developers.facebook.com/docs/messenger-platform/overview>
16. "Underthesea - Vietnamese NLP Toolkit." [Online]. Available: <https://underthesea.readthedocs.io/en/latest/readme.html>
17. "Vietnamese models for spaCy." https://github.com/trungtv/vi_spacy
18. Mady Mantha, "Introducing DIET: state-of-the-art architecture that outperforms fine-tuning BERT and is 6X faster to train," 2020. <https://rasa.com/blog/introducing-dual-intent-and-entity-transformer-diet-state-of-the-art-performance-on-a-lightweight-architecture/>