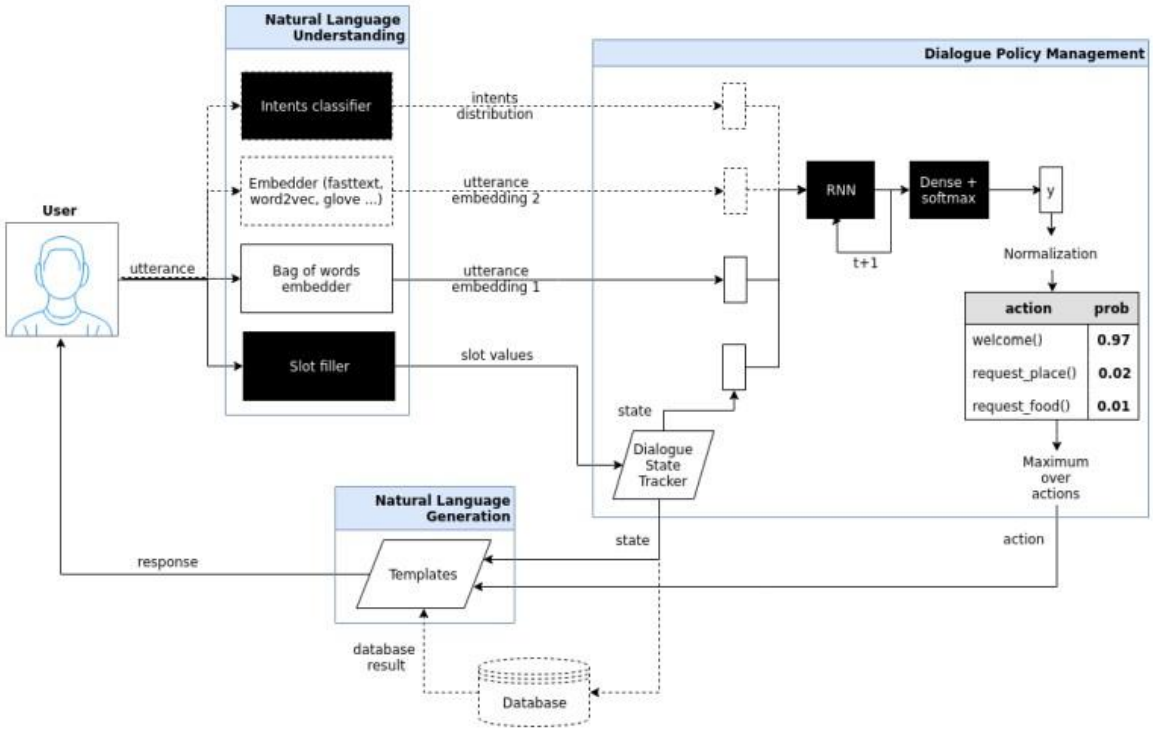
BUILDING FAQ CHATBOT USING RASA

1. Introduction

- A chatbot system uses conversational artificial intelligence (AI) technology to simulate a discussion (or a chat) with a user in natural language via messaging applications, websites, mobile apps or the telephone. It uses rule-based language applications to perform live chat functions in response to real-time user interactions.
- So many kind of chatbot that can be used in our life, but in this documentation, we will care much about a contextual chatbot- a kind of chatbot that's can make a conversation to talk about information by using the frame work rasa
- Rasa- an open source AI/framework for building chatbot. It support us to build a conversation and rule-based for our chatbot.

2. Structure of chatbot

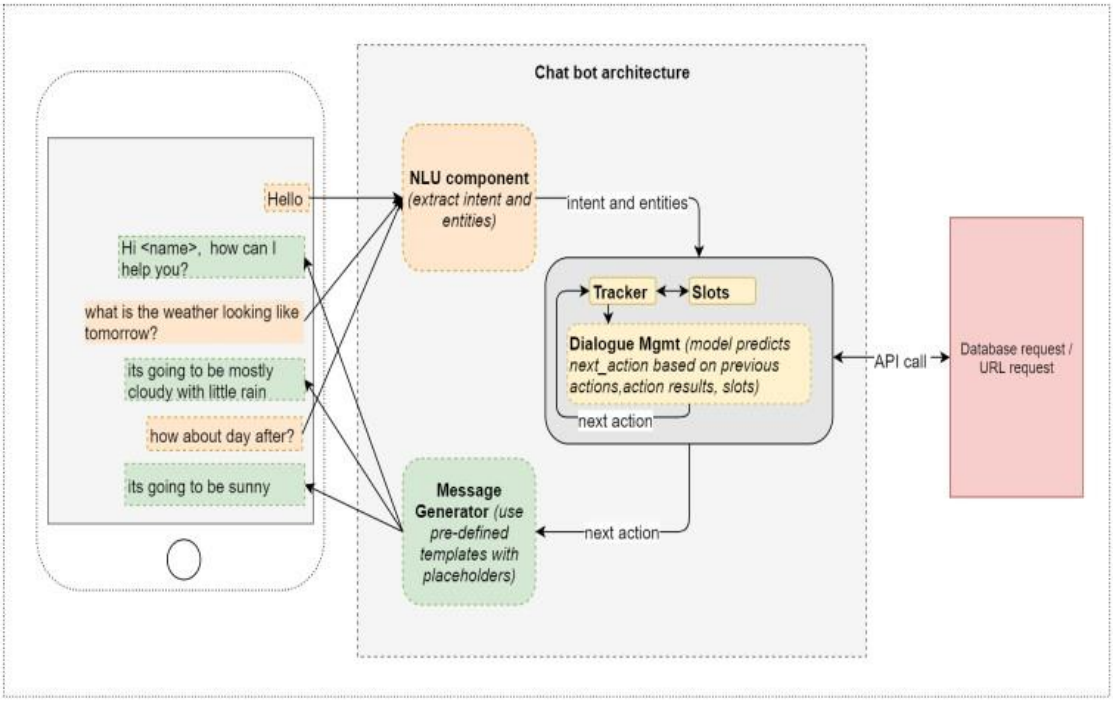


Hình 3: Cấu trúc các thành phần cơ bản hệ thống Chatbot [12]

A basic chatbot can have so many things like in picture H.1, but it must have at least 3 compulsorry components:

- Natural Language Understanding(NLU): It make chatbot understanding what we teach for our chatbot by processing natural language. More over it will extract intents and entites from user's question
- Dialogue Management(DM): Predict the action of chatbot based on previous action, action result or from intents and entites.

- Natural Language Generator(NLG): It make messages generator from policy and action that identify from DM. From that it can make a answer for the question or can based on the pre-define template that we train for our bot.
  To understanding how chatbot doing, we can take a look in Picture H.2



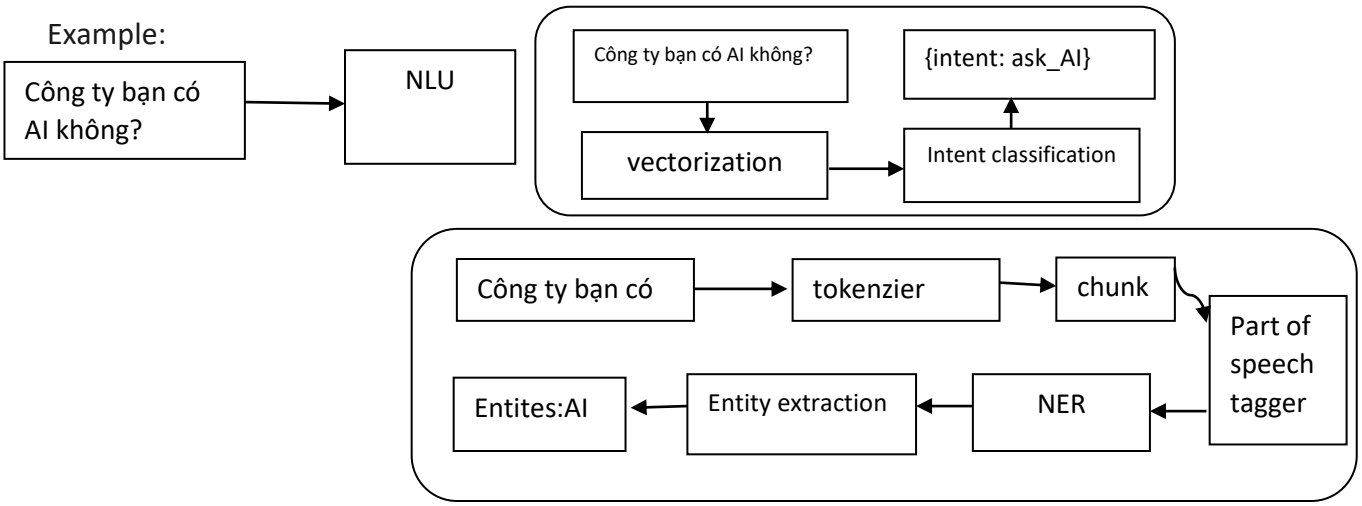
Hình 4: Mô hình các thành phần xử lý trong Chatbot [1]

2.1.1 NLU

We can say that this is the most important things for our chatbot. The smartest chatbot or not will based on this. Target of this components is that extract intents and entities from the user's saying.

- First of doing this component is that we have to get intents classification from user when they ask/say something
- The next step is extracting enties(slot filter) in user's question/saying. From this chatbot can reply the answer
- Finally, NLU handling user's messages through pipeline that sequential processing steps.

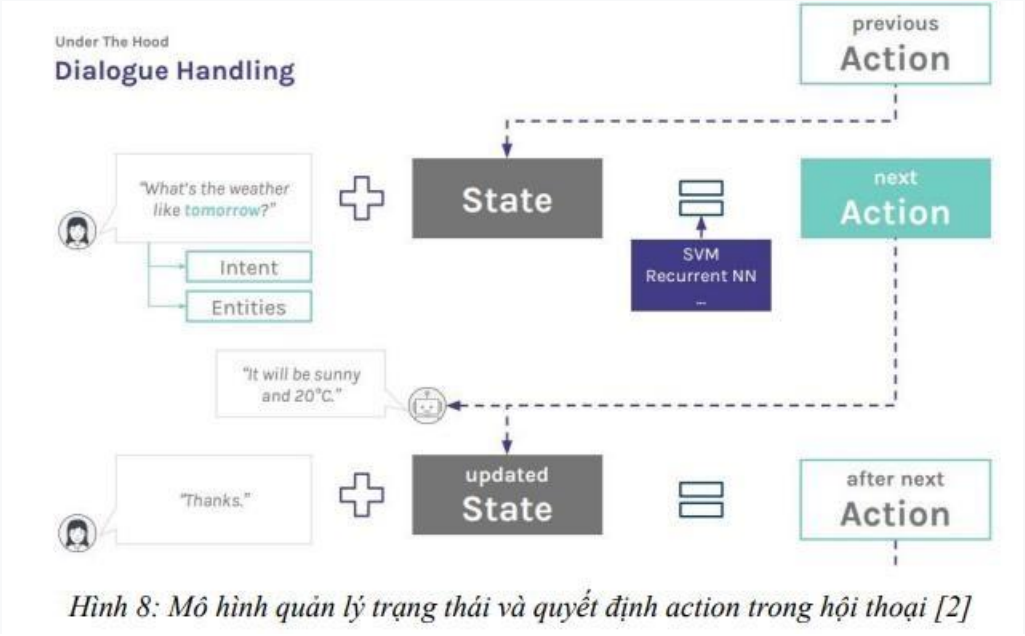
Example:



2.1.2 DM

In long conversations between Chatbots and users, Chatbots will need to remember information about context or manage conversation states. The task of conversation management plays an important role to ensure the exchange between users and Chatbot is smooth.

The task of the conversation manager is to receive input from the NLU, to manage the dialogue state(the dialogue context), and to be the important output for the language generation component (Natural Language Generation, abbreviated as NLG).

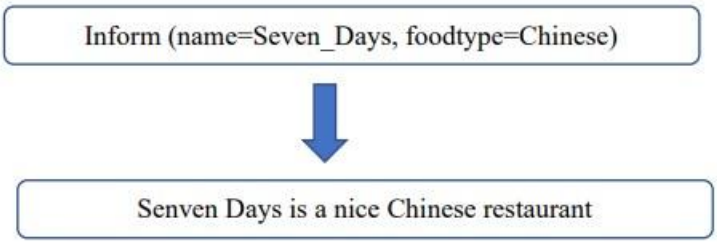


Hình 8: Mô hình quản lý trạng thái và quyết định action trong hội thoại [2]

The dialog state is saved and relies on the dialog policy set to decide the next action for the Chatbot's answer in a conversation scenario, or the action just depends on the conversation. depends on its previous dialog state.

2.1.3 NLG

NLG is Chatbot's answer generation model. It relies on mapping conversation manager actions into natural language to respond to users.



Template-based NLG:

This answer mapping method is to use predefined bot response templates to generate answers.

Semantic Frame	Natural Language
confirm()	"Please tell me more about the product your are looking for."
confirm(area=$V)	"Do you want somewhere in the $V?"
confirm(food=$V)	"Do you want a $V restaurant?"
confirm(food=$V,area=$W)	"Do you want a $V restaurant in the $W."

Hình 11: Phương pháp sinh ngôn ngữ dựa trên tập mẫu câu trả lời [1]

- **Advantages**: Simple, easy control. Suitable for closed domain problems.

- **Disadvantages**: Time consuming to define the rules, not natural in the answer. For large systems, it is difficult to control the rules, resulting in the system being difficult to develop and maintain.

3.Building chatbot using rasa framework

3.1.1. Problem

There are many application problems for Chatbot but I choose the problem for giving the information of my company to solve difficulites when customers ask something about my company, want to buy our application.

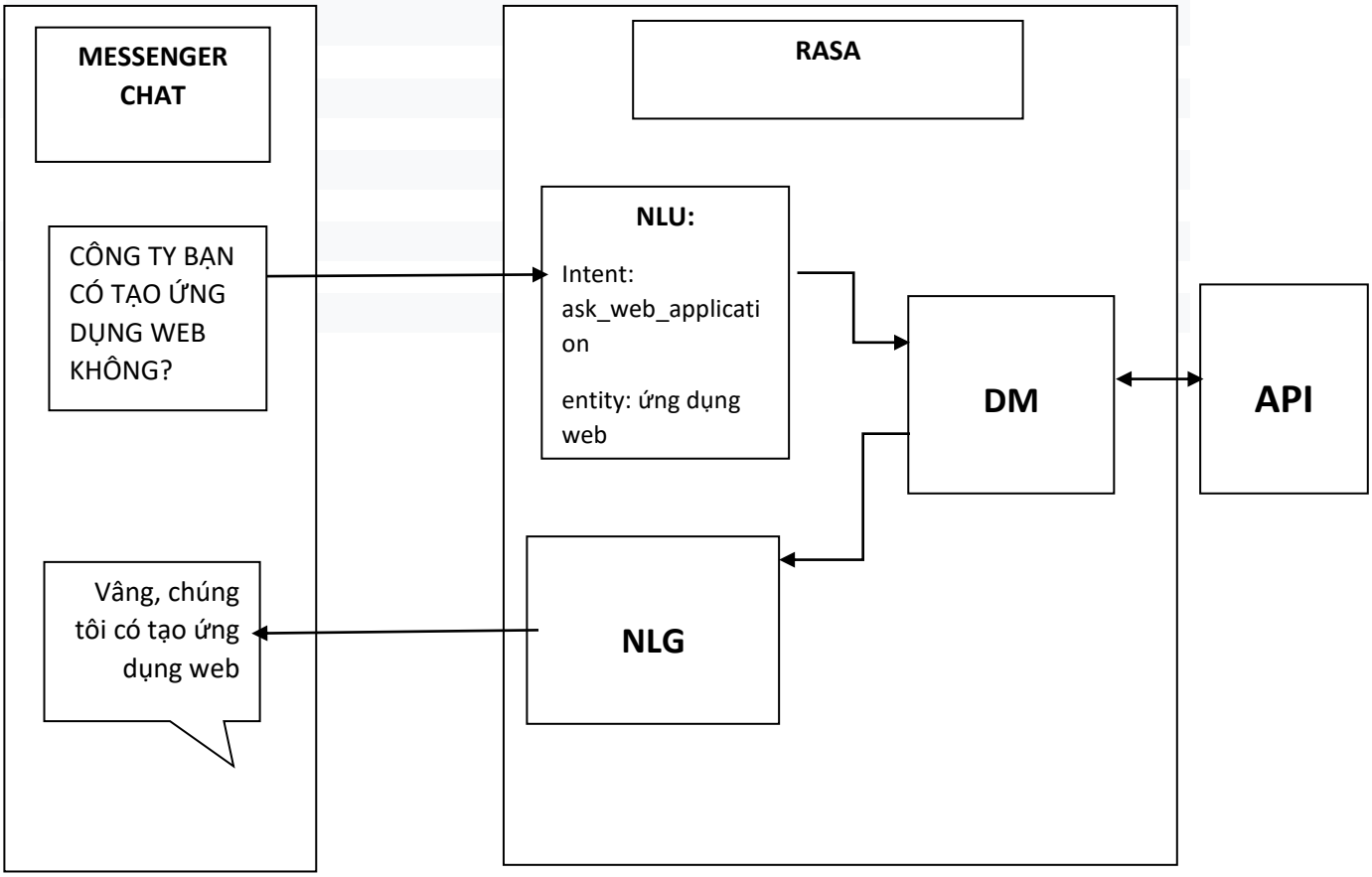
The problem that I build will focus on consulting,giving information and sales support functions

on Facebook Fanpages. The main functions include:

- Information of company
- Information of application
- Information of chatbot
- Advise of buying application

3.2.1. Building a chatbot using Rasa

Structure of chatbot using rasa will be build like this:



From the above model, the input to the system is a question user.The first is Chatbot's answer.

Components inside the system Chatbot will undertake the following:

NLU: Responsible for language vectorization, user intent classification and extract the information.

DM: Based on the status and context of the conversation to determine the action reason for the above input sentence. This component is also responsible for connecting to the base data to get data from Chatbot's response generation service system for NLG's component

NLG: Model for generating answers based on data from the DM component according to pre-built templates of sentence templates

3.3.1. Config of pretraining for our training data

Rasa provides us with 2 main methods of building training data for Chatbots:

- Pretrained Embeddings (Intent_classifier_sklearn) : Intent Classification will be based on data sets that are prefiltered, then used to represent each word in a user message as an embedded word to represent the language as a vector (word2vec). This dataset can be provided from Spacy or MITIE…

- Supervised Embeddings (EmbeddingIntentClassifier): With this method, the user will have to build the data from due to the lack of training data available. But for problems in a domain closed field, it will ensure much more accuracy and avoid data redundancy compared to the method above.

In this documentation we will build a supervised embedding to build our dataset.

```yaml
language: vi

pipeline:
# # No configuration for the NLU pipeline was provided. The following default pipeline was used to train your model.
# # If you'd like to customize it, uncomment and adjust the pipeline.
# # See https://rasa.com/docs/rasa/tuning-your-model for more information.
#   - name: WhitespaceTokenizer
#   - name: RegexFeaturizer
#   - name: LexicalSyntacticFeaturizer
#   - name: CountVectorsFeaturizer
#   - name: CountVectorsFeaturizer
#     analyzer: char_wb
#     min_ngram: 1
#     max_ngram: 4
#   - name: DIETClassifier
#     epochs: 100
#   - name: EntitySynonymMapper
#   - name: ResponseSelector
#     epochs: 100
#   - name: FallbackClassifier
#     threshold: 0.3
#     ambiguity_threshold: 0.15

# Configuration for Rasa Core.
# https://rasa.com/docs/rasa/core/policies/
policies:
# # No configuration for policies was provided. The following default policies were used to train your model.
# # If you'd like to customize them, uncomment and adjust the policies.
# # See https://rasa.com/docs/rasa/policies for more information.
#   - name: MemoizationPolicy
#   - name: RulePolicy
#   - name: UnexpecTEDIntentPolicy
#     max_history: 5
#     epochs: 100
#   - name: TEDPolicy
#     max_history: 5
#     epochs: 100
#     constrain_similarities: true
```

With the problem of focusing on a closed domain like our Chatbot and difficult to find pretrained datasets. then we don't need to use pretrained word embeddings) instead, we will create our own training dataset. This also ensures that the Chatbot has a short training time with high accuracy higher

3.4.1. Building our dataset

A conversation between a Chatbot and a customer is to solve a problem any customer request.

Example:

USER	BOT
Công ty bạn có tạo ứng dụng web không?	Vâng, chúng tôi có

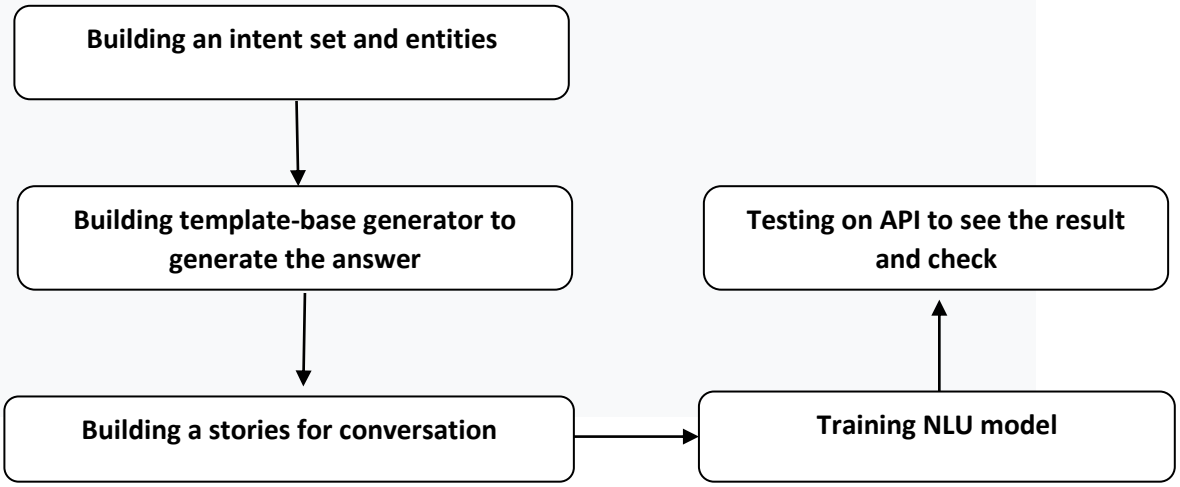
Công ty bạn đã có bao nhiêu năm kinh nghiệm trong việc tạo ứng dụng web?	Chúng tôi đã có gần 2 năm kinh nghiệm trên thị trường và hơn 3 năm trong việc tạo ứng dụng web
Các bạn đánh giá một sản phẩm ứng dụng web như nào?	Chúng tôi thường dựa vào nhiều tiêu chí đánh giá, tuy nhiên quan trọng nhất vẫn là khả năng phát triển

So with the article, build training data for Chatbot by closed domain in general, we have to build a list stories like the example above.

User's intent can be expressed in many different ways but share the same label of their message, for each question, request in the chat is a user's intent. For example users can express the purpose of their greetings by saying "xin chào ad", "ad có khỏe không?",…etc.

Once you have get the user's intent in their sentence through the conversation, we need to extract the specific information of user. Those are entities, these entities will be saved into the slot to save the context to avoid having to ask again and again in the conversation, this will serve the following actions of the Chatbot.

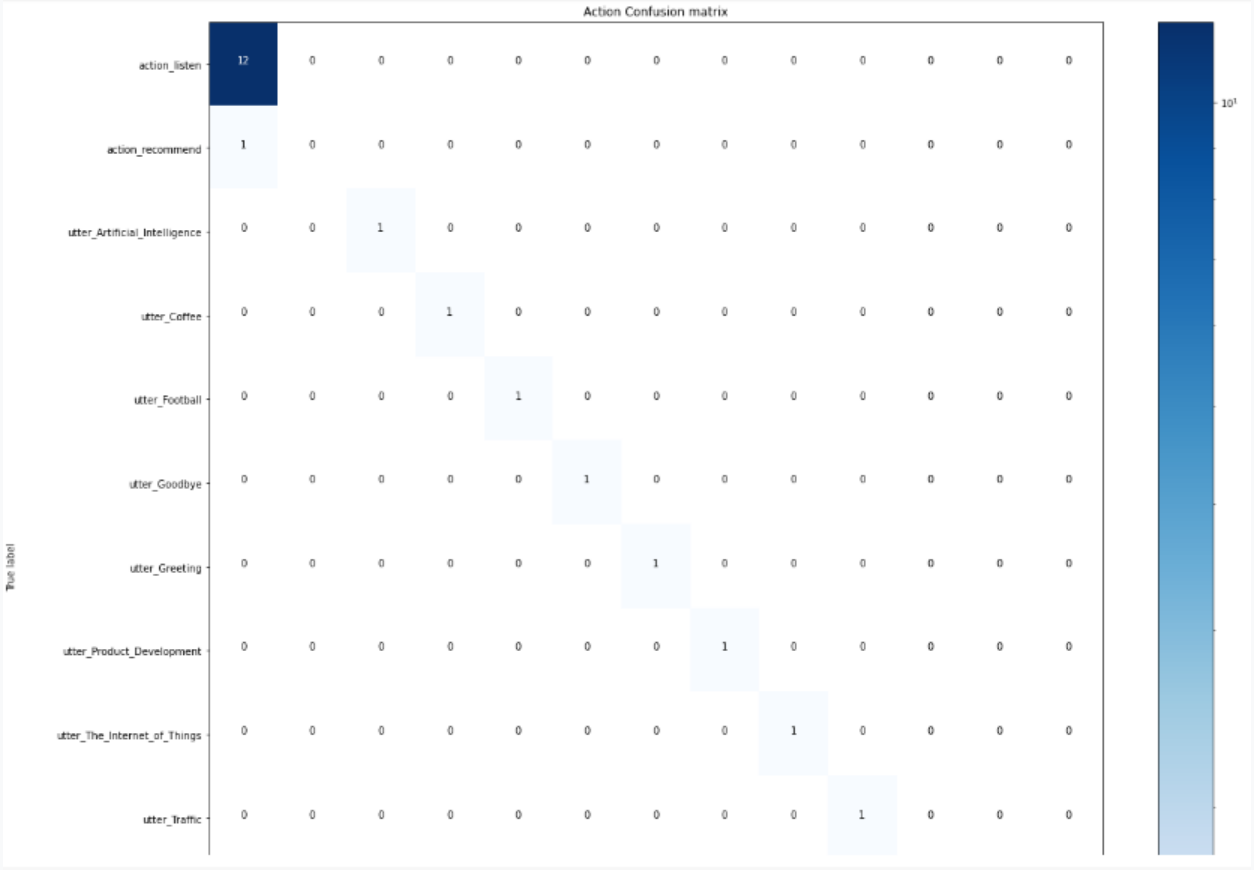
The process of building a Chatbot system can be done through the following main steps:



3.5.1. Experiments

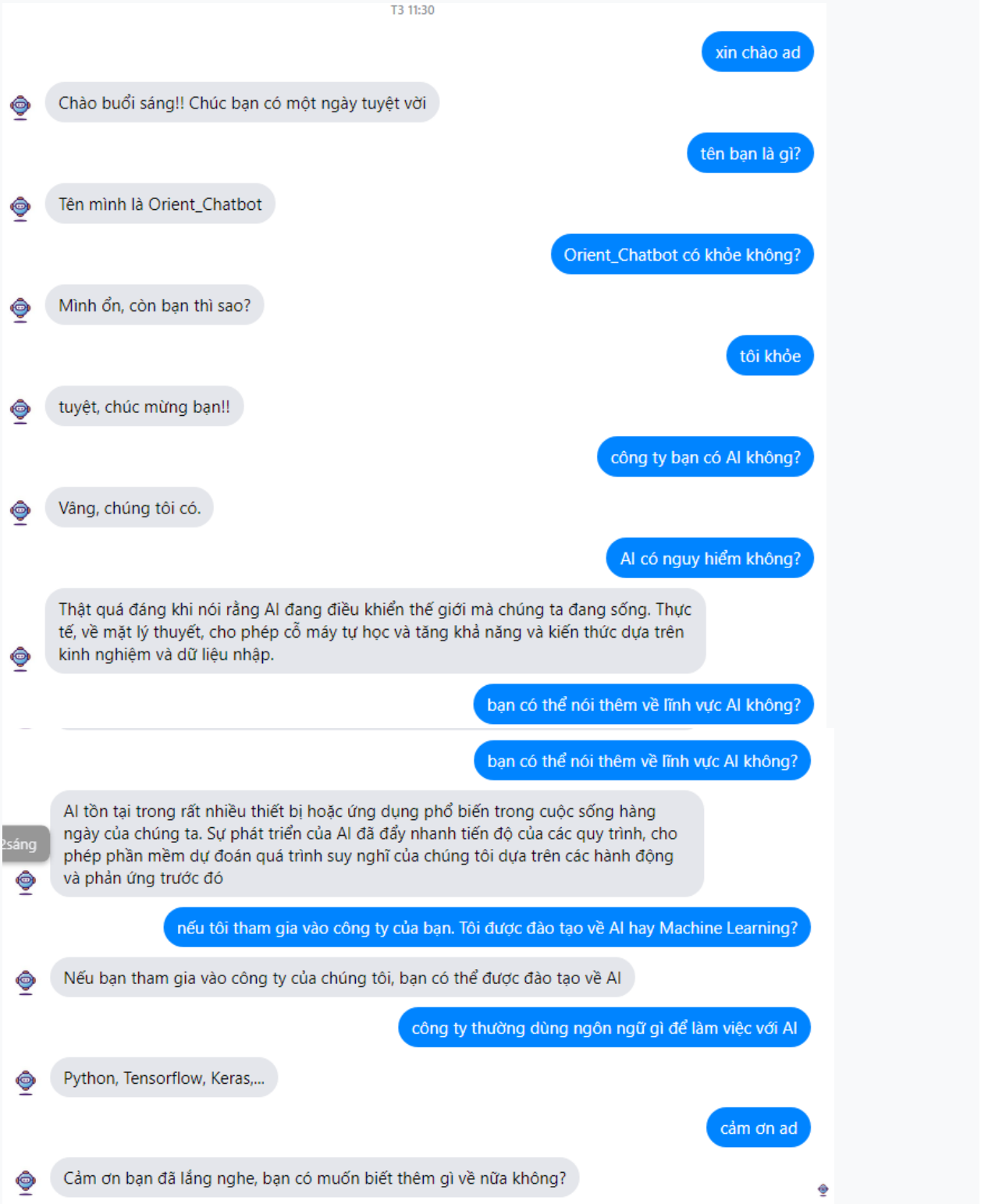
Conduct an interaction test with Chatbot with some sentences and see Chatbot's accuracy is about 80% and only counts pulse questions around the trained script for the Chatbot. As for the questions that haven't yet training(out of scopes) then the bot will give some sample navigation answers users to a number of available questions that Chatbot can answer.

Some evaluation charts for training datasets and stories (history) for bot:



With the above estimation chart, we see the data of the input sentences for Users's intent are less likely to conflict or confuse with each other. This ensures correctness of the data set for training the bot.

Demo result:


T3 11:30
xin chào ad
Chào buổi sáng!! Chúc bạn có một ngày tuyệt vời
tên bạn là gì?
Tên mình là Orient_Chatbot
Orient_Chatbot có khỏe không?
Mình ổn, còn bạn thì sao?
tôi khỏe
tuyệt, chúc mừng bạn!!
công ty bạn có AI không?
Vâng, chúng tôi có.
AI có nguy hiểm không?
Thật quá đáng khi nói rằng AI đang điều khiển thế giới mà chúng ta đang sống. Thực tế, về mặt lý thuyết, cho phép cỗ máy tự học và tăng khả năng và kiến thức dựa trên kinh nghiệm và dữ liệu nhập.
bạn có thể nói thêm về lĩnh vực AI không?
bạn có thể nói thêm về lĩnh vực AI không?
AI tồn tại trong rất nhiều thiết bị hoặc ứng dụng phổ biến trong cuộc sống hàng ngày của chúng ta. Sự phát triển của AI đã đẩy nhanh tiến độ của các quy trình, cho phép phần mềm dự đoán quá trình suy nghĩ của chúng tôi dựa trên các hành động và phản ứng trước đó
nếu tôi tham gia vào công ty của bạn. Tôi được đào tạo về AI hay Machine Learning?
Nếu bạn tham gia vào công ty của chúng tôi, bạn có thể được đào tạo về AI
công ty thường dùng ngôn ngữ gì để làm việc với AI
Python, Tensorflow, Keras,...
cảm ơn ad
Cảm ơn bạn đã lắng nghe, bạn có muốn biết thêm gì về nữa không?

As you can see that, the result of our demo in messenger's facebook, when we ask something about our company, the bot will reply the messages and it likely correct 90% as it template we have built.

4. Evaluation

From the process of learning, building and objective results from reality I have the following assessments:

- Build training data and conversation is extremely important. These are two factors that have a great influence on the flexibility and intelligence of Chatbot.
- It is also important to define and build intents meticulous analysis with different topics to be able to come up with good intention sets. For intents that are not clear or semantically close to each other will causes the decrease of accuracy of the Chatbot. So the design of intents and slots are extremely important.
- With the conversations in the pre-built history, the Chatbot responds very good when giving the correct answer to the user. However, the construction of the history for Chatbot is difficult because it can happen a lot of cases. For long conversations it is very complicated in storing slots to save conversation context
- Chatbot has the ability to randomly answer the sentence patterns in the template make the dialogue more natural. Besides the bot has the ability navigating the user to the available answer template. However, navigation is also based on the randomness of the bot leads to the fact that the bot is not smart in handling it situations like this.

5. References

[1] https://medium.com/oracledevs/say-hello-to-red-samurai-contextual-chatbot-with-tensorflow-deep-neural-network-learning-aa69dd3be04f

[2] https://rasa.com/docs/rasa/

[3] https://viblo.asia/p/hieu-rasa-qua-quy-trinh-xay-dung-mot-chatbot-giup-ban-tra-loi-cau-hoi-hom-nay-an-gi-WAyK8P4pKxX

[4] https://viblo.asia/p/tap-tanh-lam-rasa-chatbot-gAm5y8Nwldb