
Dựng ChatBot với RASA

cuong@techmaster.vn

Ứng dụng ChatBot

Nhân viên tư vấn	AI ChatBot
Chi phí lương 5-10 triệu/tháng x 3-5 người	1 triệu/tháng chi phí server. Không yêu cầu tăng lương
Trao đổi mềm mại, tự nhiên, tinh tế. Xử lý nhiều hội thoại cực phức tạp.	Chỉ trao đổi theo kịch bản huấn luyện sẵn. Kịch bản càng chi tiết trả lời càng tốt. Chưa thể tự nhiên, uyển chuyển như người
Phải đào tạo lại cho nhân viên mới từ 1 tuần đến 2 tháng	Đào tạo một lần hoặc nhiều lần, nhưng chi phí đào tạo bù xung gần như bằng 0.
Làm việc khoảng 40 giờ một tuần	Làm việc $7 * 24 = 168$ giờ Năng suất làm việc không bị ảnh hưởng bởi ngoại cảnh
Thu thập thông tin khách hàng thủ công, nhiều khi giữ riêng cho mình.	Thu thập thông tin khách hàng tự động, nhập vào một kho dữ chung, tìm kiếm được.
Cảm nhận cảm xúc khách hàng, nhưng tối kén	Có thể hiểu được cảm xúc khách hàng: hài lòng, bức xúc
Chịu áp lực kém	Chịu áp lực cực cao, chấp tất cả các thể loại tổng xỉ vả, xúc phạm, đe doạ mà vẫn bình tĩnh
Khả năng học, tiếp thu mỗi nhân viên tư vấn khác nhau, thiếu đồng nhất	Nhiều người cùng dạy chat bot mỗi ngày một ít
Thông tin khách hàng chat có thể bị rò rỉ, mất khi nhân viên tư vấn chuyển việc	Thông tin khách hàng chat được bảo toàn. Chatbot không chịu việc !

Ứng dụng Chatbot

- 63 % people would consider interacting with a chatbot (mindshareworld)
- 80 % businesses are predicted to have a chatbot by the year 2020 (Business Insider)

Du lịch, giao thông

- Hỏi thông tin chuyến bay, tuyến xe
- Nhận thông báo chuyến bay, giờ khởi hành
- Đăng ký chỗ, chọn menu ăn...

09:15	Aberdeen	BD674	Delayed to 10:55
09:45	Newcastle	BA1326	Cancelled
09:55	Glasgow	BA1476	Cancelled
09:55	Durham Tees	GF5232	Cancelled
09:55	Cork	AA8025	Delayed to 11:10
10:05	Dublin	AA7991	Delayed to 11:35
10:10	Shannon	AA8017	Delayed to 10:55
10:35	Edinburgh	BA1442	Cancelled
10:50	Manchester	BA1388	Cancelled

Hỗ trợ khách hàng, IT Help Desk

- Hỏi đáp, yêu cầu hỗ trợ các chủ đề cụ thể thay cho FAQ
- Đăng ký chỗ (cắt tóc, rửa xe, đặt bàn ăn)
- Yêu cầu tư vấn về sản phẩm
- Tư vấn triệu chứng COVID...



Khảo sát trực tuyến

Thay vì phải trả lời một bảng câu hỏi dài, khách hàng có thể trả lời hỏi đáp với bot theo kịch bản linh động



ChatBot thay thế hoàn toàn con người chưa?

- Chưa thể !
- Chatbot chỉ trả lời tốt theo kịch bản có sẵn. Kịch bản chuẩn bị càng chi tiết, khéo thì trải nghiệm người dùng sẽ tốt hơn
- Không phải cứ áp dụng chatbot là thành công. Khách hàng vẫn thích được phục vụ bởi người thật hơn



Hai nhà hàng lớn ở Trung Quốc đã buộc phải đóng cửa sau khi thay thế đội ngũ phục vụ bằng xương bằng thịt bằng các robot. Hóa ra đây lại là một sai lầm to lớn khiến họ phải trả giá đắt bằng việc đóng cửa nhà hàng.

Theo tờ *Worker's Daily* những robot đắt tiền không có khả năng nhận đơn gọi món, mang soup và rót nước. Điều đó có nghĩa là một trí thông minh nhân tạo tinh vi vẫn thua trí não của một con người.

Một vài năm trước, chủ sở hữu một nhà hàng có robot phục vụ từng tự hào nói "so với người lao động sống thì robot "không thể bệnh và xin nghỉ phép", robot vẫn phục vụ tốt như con người".

Tuy nhiên, thực tế hiện tại thì không hề như mong muốn của các chủ nhà hàng.



Thích thú với robot bồi bàn ở quán cà phê Hà Nội

Chia sẻ



Dân trí Mới đây, quán cà phê robot phục vụ đang thu hút được rất nhiều sự quan tâm, đặc biệt là trẻ em. Nhiều người tò mò cũng đến quán chỉ để tận hưởng cảm giác được robot phục vụ xem có gì khác biệt so với nhân viên bình thường.

>>Trung Quốc: Độc đáo nhà hàng do... robot phục vụ

>>Rộ mốt manocanh quay tay, hàng cháo lòng cũng mua về hút khách

>>Ông già Noel lắc hông, thổi kèn giá tiền triệu, hàng trà chanh cũng mua về hút khách



Robot thay thế con người, nghề bồi bàn liệu đang dần biến mất?

Cũng đã đóng cửa sau 3 tháng hoạt động lỗ hơn 200 triệu !

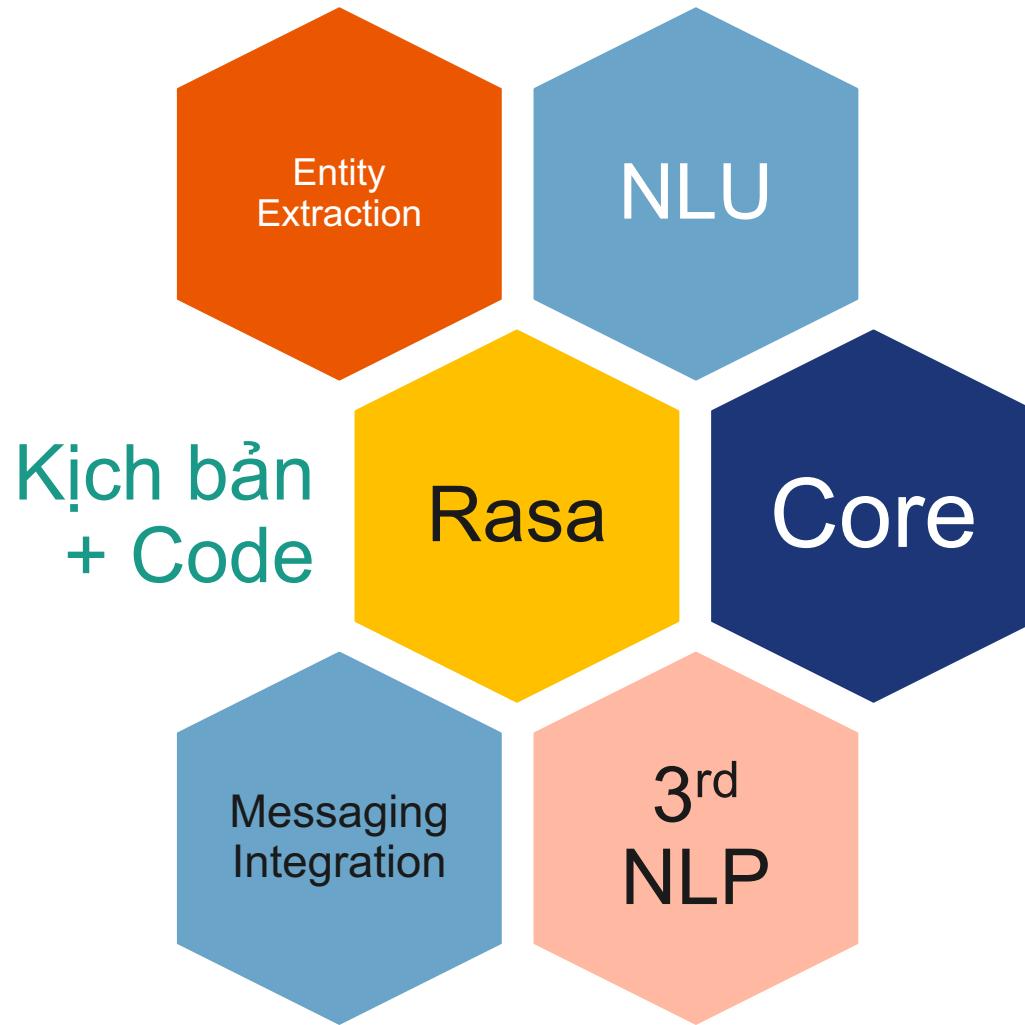
Chủ cửa hàng anh Đỗ Trung Thành, giảng viên Arduino ở Techmaster

Nên làm

- Kết hợp ChatBot với con người
- Liên tục cập nhật kiến thức Chatbot mỗi ngày
- Xem lại lịch sử trao đổi của ChatBot để cải tiến, cập nhật kịch bản
- Can thiệp ngắt ChatBot khi cần thiết

RASA





Cần dụng Chatbot từ cực đơn giản đến phức tạp

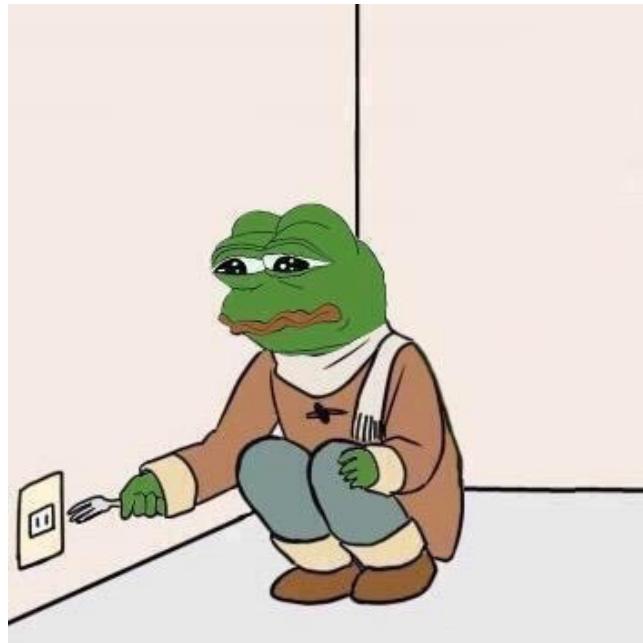


Dự án ChatBot RASA chuyên nghiệp sẽ cần

- Hội thoại phủ được càng nhiều khả năng hội thoại càng tốt
- Kịch bản phục vụ mục tiêu cụ thể
- Lấy được thông tin người dùng lưu vào trong CSDL
- Kết nối REST API để trả về dữ liệu người dùng yêu cầu
- Lưu toàn bộ hội thoại vào CSDL để phân tích trong tương lai
- Tích hợp với NLP xử lý ngôn ngữ bản địa: tiếng Việt
- Tích hợp Duckling để bóc tách nội dung
- Kết nối vào Facebook, Zalo, Messenger...
- RASA X

Cú pháp xây dựng RASA ChatBot chỉ gồm có MarkDown, YAML và Python. Nếu sai cú pháp, việc debug rất khó khăn và lâu, do đó cần quản lý mã nguồn dự án RASA bằng git:

Lưu lại trạng thái code trước đó để so sánh, quay lui nếu gặp lỗi



Cài đặt RASA trên máy tính



Thứ tự cài đặt



- **Python 3.7.4** hoặc mới hơn
- **Virtual Environment + PIP**
- **Git** quản lý phiên bản
- **Docker** để chạy Facebook Duckling bóc tách thông tin từ text người dung nhập vào
- **Visual Studio Code** với plugin soạn thảo YAML, Markdown và Python



Hướng dẫn trong slide này thực hiện
trên MacOSX 10.13.6. Các lệnh gõ trên
terminal tương đương với gõ trên
terminal trên Ubuntu hay Windows

```
$ python3 -version
```

Kiểm tra phiên bản

```
Python 3.7.x
```

```
$ mkdir hellobot
```

Tạo thư mục chứa dự án

```
$ cd hellobot
```

Chuyển vào bên trong thư mục

```
$ python3 -m venv venv
```

Tạo môi trường ảo Python

```
$ source venv/bin/activate
```

Kích hoạt

```
$ pip install - requirements.txt
```

rasa
spacy

requirements.txt

File này liệt kê các Python module cần phải cài đặt

Gõ thêm lệnh này để tại Spacy. Spacy là thư viện xử lý ngôn ngữ tự nhiên, sẽ dùng để bóc tách entity

```
$ python -m spacy download en_core_web_md\npython -m  
spacy link en_core_web_md en
```

<https://spacy.io/usage/models>

```
📁 .../hellobot ➜ 🐱 master ① ? ✓ ➔ rasa shell
```

```
2020-04-04 15:29:00 INFO    root - Connecting to channel 'cmdline' which was specified by the '--connector' argument. Any other channels will be ignored. To connect to all given channels, omit the '--connector' argument.
```

```
2020-04-04 15:29:00 INFO    root - Starting Rasa server on http://localhost:5005
```

```
Bot loaded. Type a message and press enter (use '/stop' to exit):
```

```
Your input -> Hi
```

```
Hey! How are you?
```

```
Your input -> Fine
```

```
Great, carry on!
```

```
Your input -> Are you bot
```

```
I am a bot, powered by Rasa.
```

```
Your input -> Cool
```

```
Great, carry on!
```

```
Your input -> What is your name?
```

```
Bye
```

```
Your input -> |
```

`rasa init --no prompt`

Khởi tạo dự án rasa

`rasa train`

Huấn luyện

`rasa shell`

Dùng thử dạng dòng lệnh

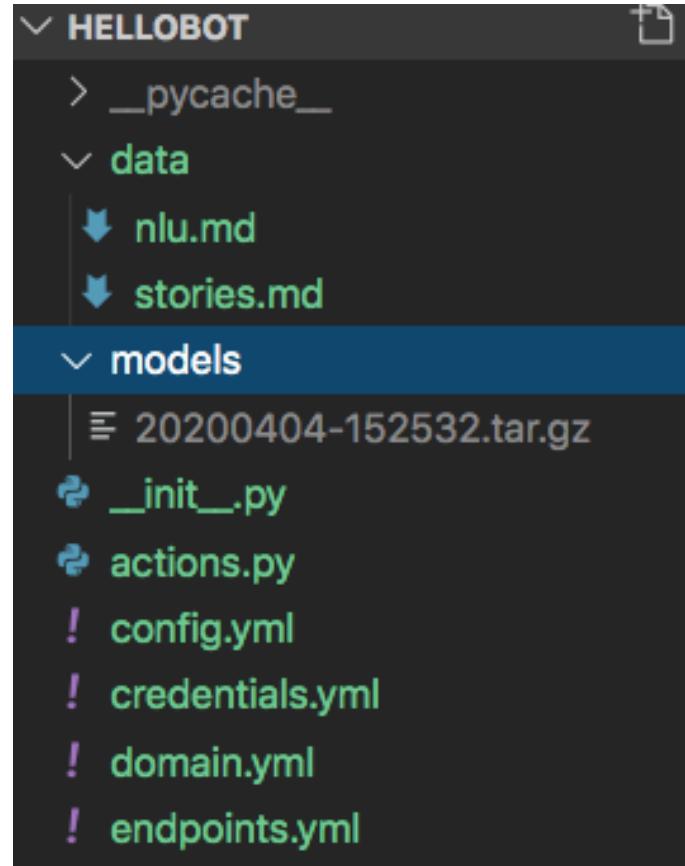
`rasa shell -debug`

Dùng thử + gỡ rối

`rasa data validate`

Kiểm tra tính hợp lệ dữ liệu

Khung dự án RASA



config.yml cấu hình pipeline, policy

endpoints.yml cấu hình đầu nối RASA
ra ngoài: action server, tracker store

credentials.yml lưu thông tin đăng
nhập Facebook, Slack, MatterMost..

config.yml cấu hình cho dự án rasa

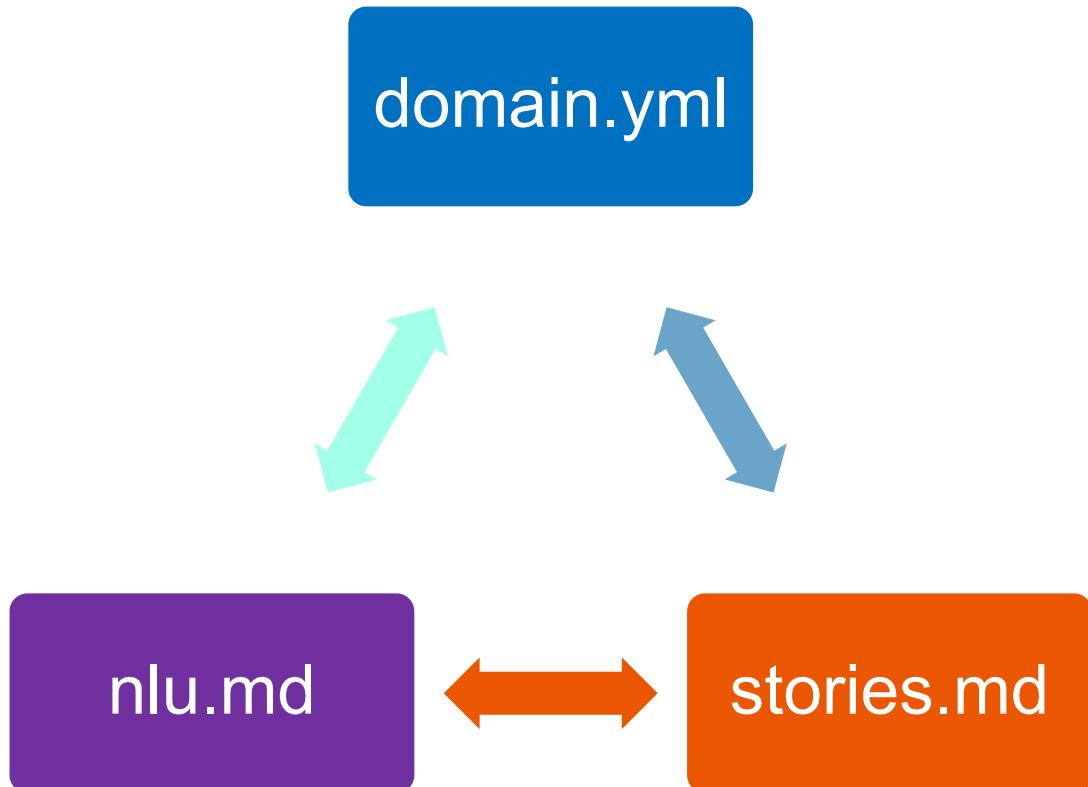
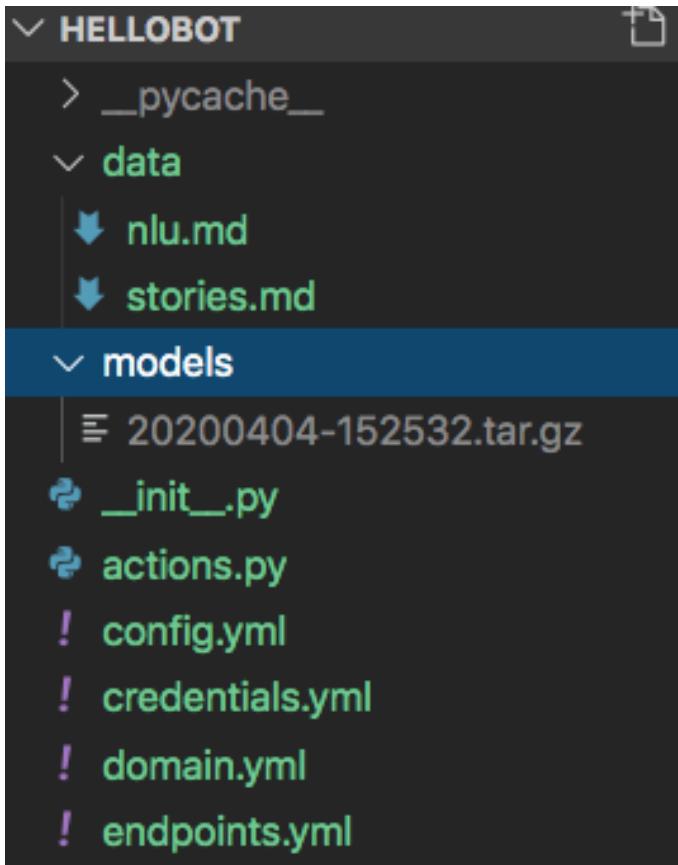
- **language:** cần xác định 1 ngôn ngữ cụ thể để việc huấn luyện tốt hơn. en: English, vi: Vietnamese
- **pipeline:** gồm các thành phần xử lý, bóc tách từ:
 - WhiteSpaceTokenizer: tách từ theo dấu cách
 - SpacyToktenizer: tách từ dùng thư viện Spacy
 - RegexFeaturizer: dung regex để phát hiện
 - Duckling
- **policies:** thiết lập hành vi của rasa
 - MemoizationPolicy: nhớ được bao nhiêu bứ
 - TEDPolicy: số epoch để huấn luyện
 - FallbackPolicy: nếu bot không hiểu thì sẽ làm gì
 - FormPolicy:





Q: Làm sao để hiểu rõ từng component và policy?

A: Hãy google để tìm hiểu, bổ sung từng component vào pipeline thí nghiệm xem rasa hoạt động thế nào.



intent:greet

- hey
- hello
- hi
- good morning
- good evening
- hey there

intent:goodbye

- bye
- goodbye
- see you around
- see you later

intent:affirm

- yes
- indeed
- of course
- that sounds good
- correct

nlu.md

intent: ý đồ, mong muốn

Khách hàng hỏi, bot trả lời

intent:greet

intent:

intent:affirm



intent;reject

Viết sai cú pháp xảy ra 2 khả năng:

1. rasa train báo lỗi, có thể sửa được ngay
2. luồng hội thoại không như ý, gỡ rối rất khó

`## intent: muon_dong_tien_mat`

- đóng tiền mặt
- trả tiền mặt
- đóng trực tiếp
- trả trực tiếp
- nhận tiền mặt không
- đến đâu đóng tiền
- muốn đóng tận nơi
- khi đi học sẽ nộp
- nộp học phí
- đến trung tâm nộp
- đến lớp nộp

nlu.md

`## intent: muon_dang_ky_hoc`

- muốn đăng ký học
- muốn học khoá này
- muốn đăng ký lớp này
- muốn học lớp này
- quan tâm khoá này
- quan tâm lớp này
- xin thông tin khoá học
- xin thông tin lớp

- Không viết câu quá dài
- Tập trung keyword khác biệt để bắt được ý đồ khách hàng
- Liệt kê càng nhiều tình huống càng tốt
- Một đoạn text không nên lặp lại ở 2 intent khác nhau



Này em ơi cho anh hỏi khi nào Techmaster
khai giảng lớp Deep Learning ấy nhỉ

Mình muốn hỏi lịch khai giảng lớp Deep
Learning



Khi nào khai giảng Deep
Learning

Deep Learning khi nào khai
giảng

Deep Learning lịch khai giảng

lịch khai giảng Deep Learning

Deep Learning lịch mở lớp

muon_dang_ky_hoc

- * muon_dang_ky_hoc
- utter_huong_dan_dang_ky_hoc
- utter_yeu_cau_goi_so_hotline

hoi_cach_mua_khoa_hoc

- * hoi_cach_mua_khoa_hoc
- utter_yeu_cau_dang_ky_chuyen_khoan
- utter_cho_so_tai_khoan_ngan_hang

muon_dong_tien_mat

- * muon_dong_tien_mat
- utter_tra_loi_khach_muon_dong_tien_mat
- utter_cho_so_tai_khoan_ngan_hang

Stories.md

Kịch bản hội thoại

- Đơn giản: hỏi đáp
- Khó hơn: chạy các action
đặt câu hỏi, lấy thông tin
từ khách hàng

```
## hoi_gio
* hoi_gio
- action_hoi_gio

## hoi_ngay
* hoi_ngay
- action_hoi_ngay
```

stories.md

```
class Action_Hoi_Gio(Action):
    def name(self) -> Text:
        return "action_hoi_gio"

    def run(self, dispatcher: CollectingDispatcher,
            tracker: Tracker,
            domain: Dict[Text, Any]) -> List[Dict[Text, Any]]:
        currentDT = datetime.datetime.now()
        dispatcher.utter_message(text = "Bây giờ là " +
currentDT.strftime("%I:%M:%S %p"))

        return []
```

actions.py

Your input -> chào em
Em chào anh/chị
Your input -> mấy giờ rồi em
Bây giờ là 05:29:09 PM



intents:

- greet
- goodbye

responses:

utter_greet:

- text: "Hey! How are you?"

utter_goodbye:

- text: "Bye"

session_config:

session_expiration_time: 60

carry_over_slots_to_new_session:
true

domain.yml

intents: Liệt kê đủ các intents, actions

responses: Viết lời đối đáp khách hàng, utter

session_config: thời lượng hội thoại tối đa 60 phút, có nhớ thông tin thu thập trước đó không

utter_phan_ung_bi_khach_chui:

- text: "Anh/chị bình tĩnh được không. Em hiểu hết các ngôn từ đó đấy ạ"
- text: "Hình như hôm nay trời nóng, anh/chị đang cáu giận thì phải"
- text: "Yêu mọi người, tin vài người, đừng xúc phạm tới ai hết"

utter_phan_ung_ham_doa:

- text: "Anh/chị bình tĩnh. Nội dung chat đã được ghi lại đấy ạ. Nếu có gì em chưa phải mong anh chị thứ lỗi"
- text: "Ít ai thấy sự tài giỏi ở người đã xúc phạm mình"
- text: "Có những ngôn từ còn tồi tệ hơn chửi rủa, có những ngôn từ đem lại sự tổn thương"

Một utter có thể chứa vài khả năng
để đa dạng hóa hội thoại



Thực hành 1: Xây dựng bot trả lời hỏi đáp theo kịch bản có sẵn

Bot Vietnam Airline

- Một ngày có bao chuyến bay Hà nội – Sài Gòn?
- Giá vé hạng phổ thông bao nhiêu?
- Hành khách được mang bao nhiêu cân (kg)?
- Có được mang nước hoa lên máy bay?
- Mang theo chó thì gửi thế nào?
- Nếu mang dao thì gửi thế nào?
- Thai phụ có được đi máy bay?
- Có xe bus đưa đón từ Hà nội đến sân bay Nội Bài không?
- Giờ khởi hành chuyến bay VNSG-178 là mấy giờ
- Có được check in trước qua bot không?

Vài kinh nghiệm xử lý lỗi khi train rasa

- Dùng <http://yamllint.com> để kiểm tra domain.yml có đúng cú pháp không?
- Kiểm tra cú pháp markdown ở các file nlu.md và stories.md

```
## chao
* chao
  - utter_chao
  -
```

Khai báo thiếu

rasa data validate

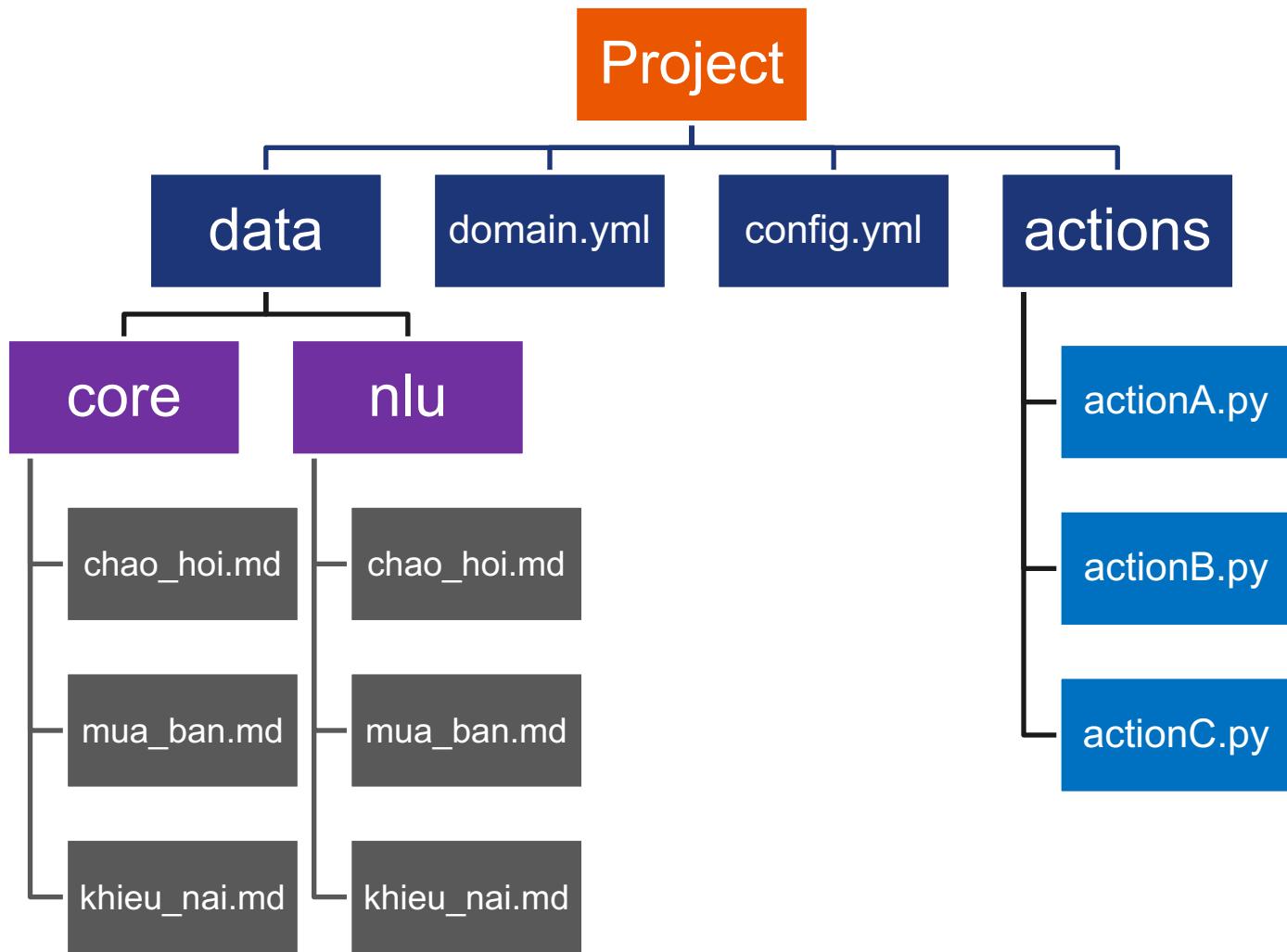
```
📁 .../demo ➜ 🐱 master ! ? ➔ ✅ ➔ rasa data validate
2020-04-04 23:16:25 INFO    rasa.validator - Validating intents...
/Users/techmaster/techrasha/venv/lib/python3.7/site-packages/rasa/utils/common.py:347: UserWarning: There
is a message in the training data labelled with intent 'cam_ket_viec_lam'. This intent is not listed in y
our domain. You should need to add that intent to your domain file!
More info at https://rasa.com/docs/rasa/core/domains/
2020-04-04 23:16:25 INFO    rasa.validator - Validating uniqueness of intents and stories...
/Users/techmaster/techrasha/venv/lib/python3.7/site-packages/rasa/utils/common.py:347: UserWarning: The ex
ample 'chưa' was found labeled with multiple different intents in the training data. Each annotated messa
ge should only appear with one intent. You should fix that conflict The example is labeled with: nguoi_ba
t_dau, tu_choi.
/Users/techmaster/techrasha/venv/lib/python3.7/site-packages/rasa/utils/common.py:347: UserWarning: The ex
ample 'học từ đâu' was found labeled with multiple different intents in the training data. Each annotated
message should only appear with one intent. You should fix that conflict The example is labeled with: lo_
lang_hoc_trai_nganh, nguoi_bat_dau.
/Users/techmaster/techrasha/venv/lib/python3.7/site-packages/rasa/utils/common.py:347: UserWarning: The ex
ample 'con phò' was found labeled with multiple different intents in the training data. Each annotated me
ssage should only appear with one intent. You should fix that conflict The example is labeled with: che_b
ot_ngu, chui_bot.
2020-04-04 23:16:25 INFO    rasa.validator - Validating utterances...
Project validation completed with errors.
```

Tái cấu trúc mã nguồn để bảo trì tốt hơn

- Chatbot phức tạp xử lý hàng trăm -> 1000 intent của khách hàng, việc nhập kịch bản càng ngày trở nên khó khăn, dễ lỗi;
 - Nhập lại intent đã có
 - Cùng một đoạn text có trong nhiều intent khác nhau
- Nhiều người cùng tham gia thiết kế kịch bản, sửa chõng chéo lên một file nlu.md và stories.md



```
✓ data
  ✓ core
    ↴ chao_hoi_thong_thuong.md
    ↴ hoi_chung_chung.md
    ↴ hoi_khoa_hoc.md
    ↴ hoi_trung_tam.md
    ↴ lich_hoc_khai_giang.md
    ↴ mua_khoa_hoc.md
    ↴ tan_tinh.md
    ↴ tu_van_chung.md
    ↴ tu_van_nghe.md
    ↴ xuc_pham.md
  ✓ nlu
    ↴ chao_hoi_thong_thuong.md
    ↴ hoi_chung_chung.md
    ↴ hoi_khoa_hoc.md
    ↴ hoi_trung_tam.md
    ↴ lich_hoc_khai_giang.md
    ↴ mua_khoa_hoc.md
    ↴ synonym.md
    ↴ tan_tinh.md
    ↴ tu_van_chung.md
    ↴ tu_van_nghe.md
    ↴ xuc_pham.md
```



```
## tán tỉnh bot
- nghi_la_bot
- khen_bot
- hoi_ten_bot
- hoi_tuoi
- hoi_que_quan
- hoi_co_nguo_i_yeu_chua
- hoi_nha_o_dau
- hoi_gioi_tinh
- xin_so_dien_thoai
- hoi_bot_hoc_o_dau
- muon_gap_em
- to_tinh
## xúc phạm bot
- che_bot_ngu
- chui_bot
- ham_doa_bot
```



Nhóm các intent theo category để dễ quan sát, tìm, thêm bớt

Custom Action

Tình huống

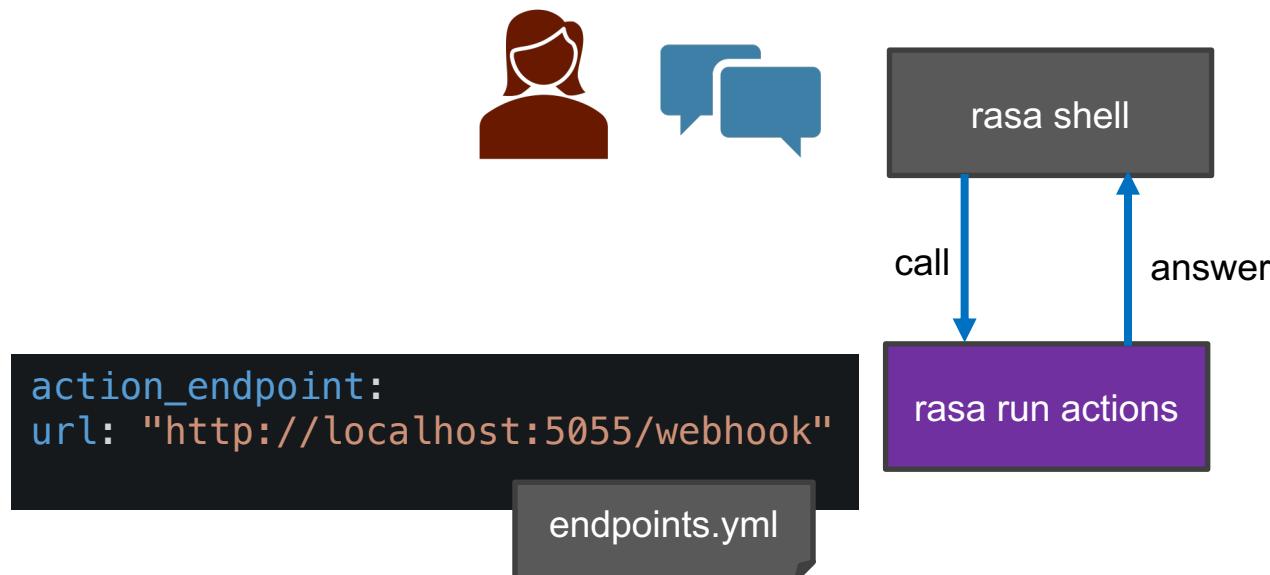
- Khách hàng hỏi giờ, hỏi thời tiết, giờ khởi hành, giá ngoại tệ
- Giám đốc hỏi doanh thu trong ngày, số lượng hàng bán trong tháng...
- Nhân viên đăng ký giữ chỗ phòng họp

Chat Bot không thể chỉ đáp lại đoạn text có sẵn mà cần phải kết nối vào cơ sở dữ liệu, REST...

Rasa custom actions sẽ được chạy ở một server riêng mặc định là **5055**

Cấu hình này được lưu trong file **endpoints.yml**

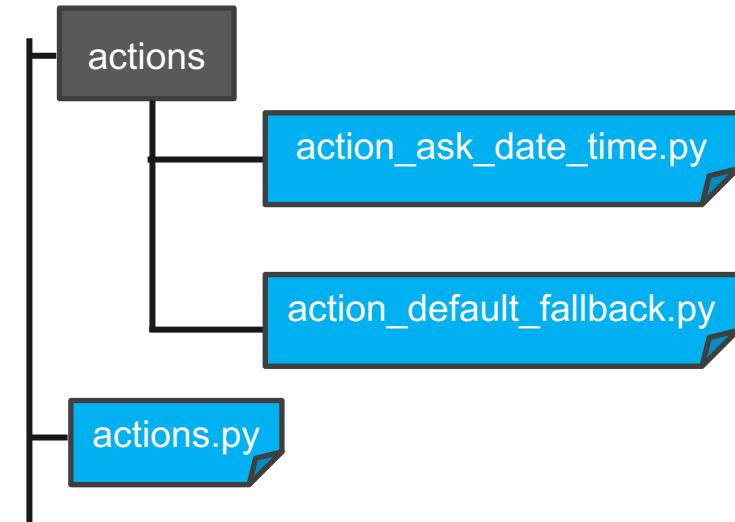
Bạn có thể đổi **action_endpoint** sang một địa chỉ IP khác, cổng khác



RASA cho phép viết Custom Action

- Khi có ít action, bạn lập trình Python trực tiếp vào file actions.py
- Nếu có nhiều actions hoặc code dài hãy tách ra thành nhiều file py rồi ở file actions.py import lại

```
from actions import action_ask_datetime  
from actions import action_default_fallback
```



```
from typing import Any, Text, Dict, List
from rasa_sdk import Action, Tracker
from rasa_sdk.executor import CollectingDispatcher
import datetime

class Action_Hoi_Ngay(Action):      Trả về tên action
    def name(self) -> Text:
        return "action_hoi_ngay" ← Logic chính của action
    def run(self, dispatcher: CollectingDispatcher,
            tracker: Tracker,
            domain: Dict[Text, Any]) -> List[Dict[Text, Any]]:
        currentDT = datetime.datetime.now()
        dispatcher.utter_message(text = "Hôm nay là " +
currentDT.strftime("%a, %b %d, %Y"))
        return []
```



Thực hành 2: Viết hàm trả lời ngày giờ

Kịch bản

- Em ơi bây giờ mấy giờ rồi / Giờ là mấy giờ nhỉ ?
 - Bây giờ là 8:34AM anh/chị ạ
-
- Hôm nay là ngày mấy nhỉ?
 - Hôm nay là thứ 4, ngày 12 tháng 3 2020
-
- Hôm này là ngày mấy trong năm?
 - Hôm này là ngày thứ 134 trong năm 2020

Lấy thông tin nhập từ người dùng

Link tham khảo

<https://blog.rasa.com/building-contextual-assistants-with-rasa-formaction/>

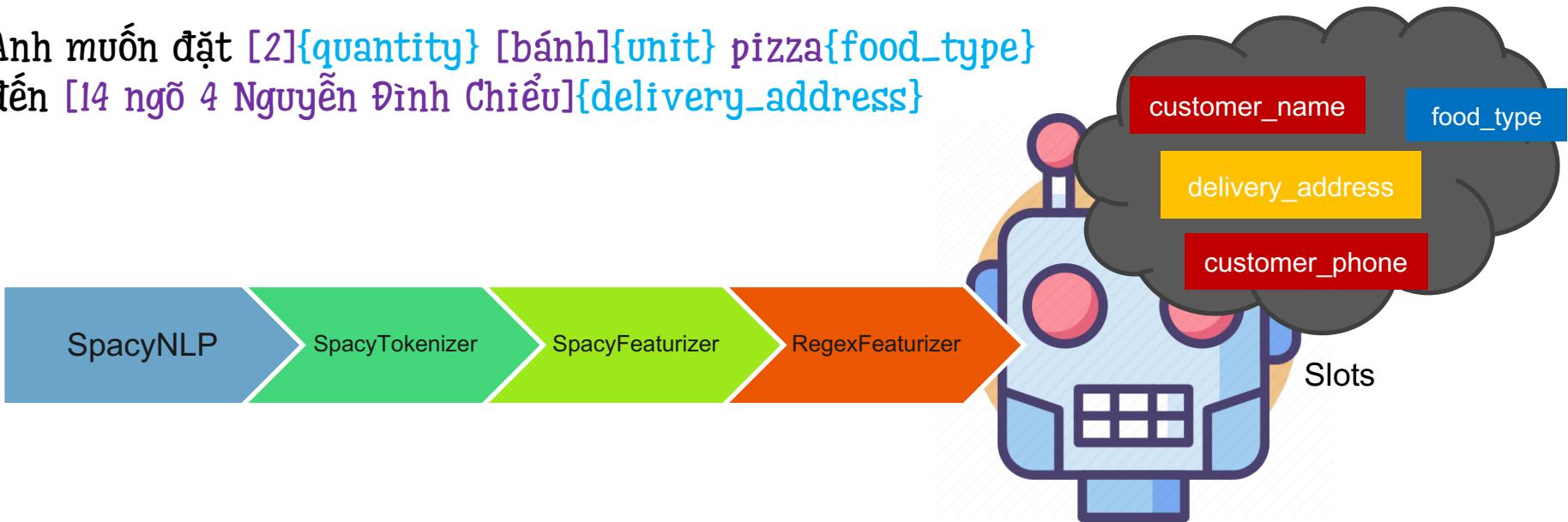
Mã nguồn của ví dụ trong tutorial này đã được nâng cấp lên phiên bản mới nhất

<https://medium.com/the-research-nest/building-yuki-a-level-3-conversational-ai-using-rasa-1-0-and-python-493e163c7911>

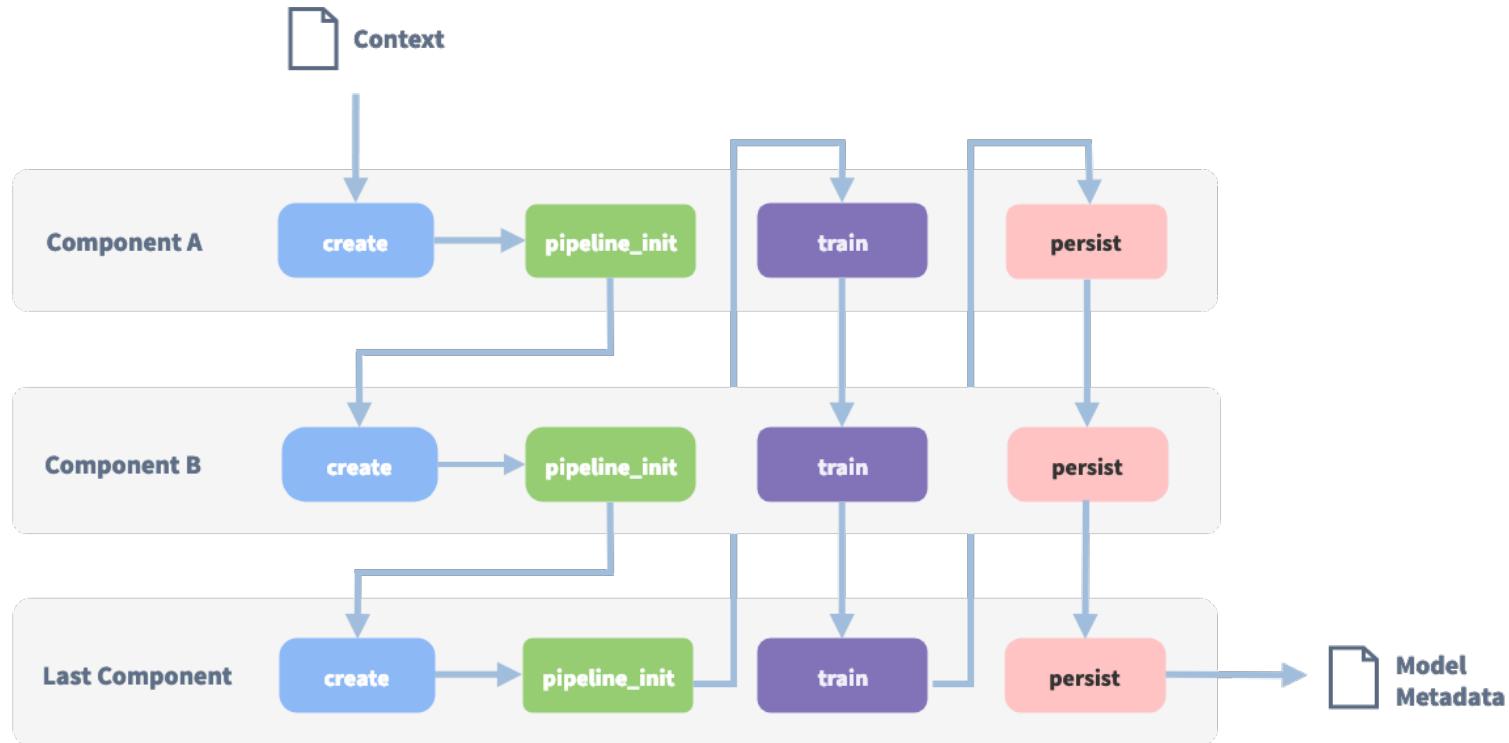
Entity khác gì Slot

- Slot là dữ liệu bot ghi nhớ
- Entity là những mẫu dữ liệu trong đoạn chat của khách hàng
- Text chat của khách hàng qua một quá trình xử lý (pipeline)

Anh muốn đặt [2]{quantity} [bánh]{unit} pizza{food_type}
đến [14 ngõ 4 Nguyễn Đình Chiểu]{delivery_address}



Component Lifecycle (train)



Ứng dụng get_news

- KH: chat với rasa yêu cầu cập nhật tin tức
- Bot: hỏi chủ đề → topic_news
- KH: gõ chủ đề → topic_news
- Bot: bóc tách topic_news sau đó kết nối vào REST API
https://newsapi.org/v2/everything?q=topic_news
Trả về kết quả

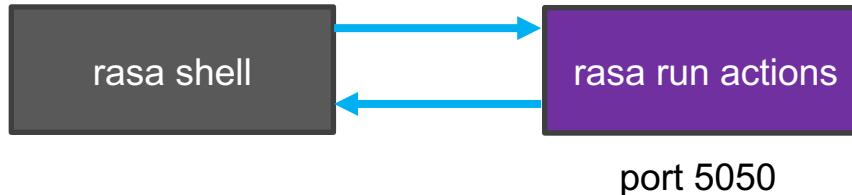
Chạy thử ứng dụng get_news

Tạo cửa sổ terminal #1

```
$ source venv/bin/activate  
$ cd get_news  
$ rasa validate data  
$ rasa train  
$ rasa shell
```

Tạo cửa sổ terminal #2

```
$ source venv/bin/activate  
$ cd get_news  
$ rasa run actions
```



Your input -> hi

Hi there! Watashi wa Yuki, as in snow in japanese.

Your input -> get me news

Give me a topic/keyword on which you would like to know the latest updates in.

Your input -> corona virus

2020-04-06 16:33:41 ERROR rasa.core.actions.action - Failed to validate slot topic_news with action get_news

Sorry, I did not understand you, please try input something else.

Your input -> get me news

Here is some news I found!

How to get the best information on coronavirus

<https://www.cnn.com/2020/03/17/opinions/coronavirus-media-literacy-matters-kirby/index.html>

How to 'Force' a Chronological Feed in Facebook's New Design

<https://lifehacker.com/how-to-force-a-chronological-feed-in-facebooks-new-desi-1842426847>

What it's like to live with someone who still doesn't take coronavirus seriously

<https://mashable.com/article/what-to-do-if-partner-friends-family-roommates-not-taking-coronavirus-seriously/>

I hope you found what you were looking for!

Your input -> get me updates sports

Here is some news I found!

How to Watch Sports During the Coronavirus Shutdown

<https://lifehacker.com/how-to-watch-sports-during-the-coronavirus-shutdown-1842377742>

Các bước để soi mã nguồn



```
entities:  
- topic_news  
forms:  
- get_news  
slots:  
    requested_slot:  
        type: unfeaturized  
topic_news:  
    type: text
```

domain.yml

intent:getNews

- Send me latest news updates
- I want to read some news
- give me current affairs
- some current affairs pls Ánh xạ từ khoá vào entity
- Find some interesting news
- News please
- Get me latest updates in [science](topic_news)
- latest updates in [sports](topic_news)
- whats the latest news in [business](topic_news)
- send news updates
- Fetch some news
- get news
- whats happening in this world
- tell me something about whats happening around
- interesting news pls
- latest updates in [blockchain](topic_news)
- get me latest updates in [astronomy](topic_news)



Nếu khách hàng gõ đúng từ khoá RASA sẽ bắt được và bóc tách



Việc hiểu người dung gõ gì cực khó !

Do đó cần phải thiết kế kịch bản + NLP dùng để hỗ trợ rasa hiểu tốt:

1. Synonym: từ đồng nghĩa
2. Lookup table: bảng tra từ
3. Entity extractor: dùng thư viện NLP
 - o SpacyEntityExtractor
 - o Duckling

<https://rasa.com/docs/rasa/nlu/entity-extraction/>

Component	Requires	Model	Notes
CRFEntityExtractor	sklearn-crfsuite	conditional random field	good for training custom entities
SpacyEntityExtractor	spaCy	averaged perceptron	provides pre-trained entities
DucklingHTTPExtractor	running duckling	context-free grammar	provides pre-trained entities
MitieEntityExtractor	MITIE	structured SVM	good for training custom entities
EntitySynonymMapper	existing entities	N/A	maps known synonyms
DIETClassifier		conditional random field on top of a transformer	good for training custom entities

Symnonym – từ đồng nghĩa

Giúp Rasa hiểu tốt được nhiều từ khác nhau nhưng đồng nghĩa

- Hà nội
- hanoi
- ha_noi
- HN
- thủ đô

Lookup table

Dùng để bắt /gom các keyword thuộc cùng một loại entity

lookup:weekday

- monday
- tuesday
- wednesday
- thursday
- friday
- saturday
- sunday

lookup:district

- hai bà trưng
- hoàn kiếm
- đống đa
- ba đình
- thanh xuân
- cầu giấy
- tây hồ

Rasa bản cũ khai báo lookup trực tiếp trong nlu.md

lookup:topic_news

data/test/lookup_tables/topic_news.txt

Rasa bản mới khai báo lookup trong file riêng biệt

```
language: en
pipeline:
- name: SpacyNLP
- name: SpacyTokenizer
- name: SpacyFeaturizer
- name: RegexFeaturizer
- name: SpacyEntityExtractor
- 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
```

<https://towardsdatascience.com/a-beginners-guide-to-rasa-nlu-for-intent-classification-and-named-entity-recognition-a4f0f76b2a96>