

Triển khai Chatbot ứng dụng vào thực tế

Đào Quang Duy

Rabiloo R&D Team





Agenda

- Giới thiệu về Chatbot
- Case Study thực tế
- Kiến trúc hệ thống
- Các công cụ hỗ trợ
- Hỏi đáp





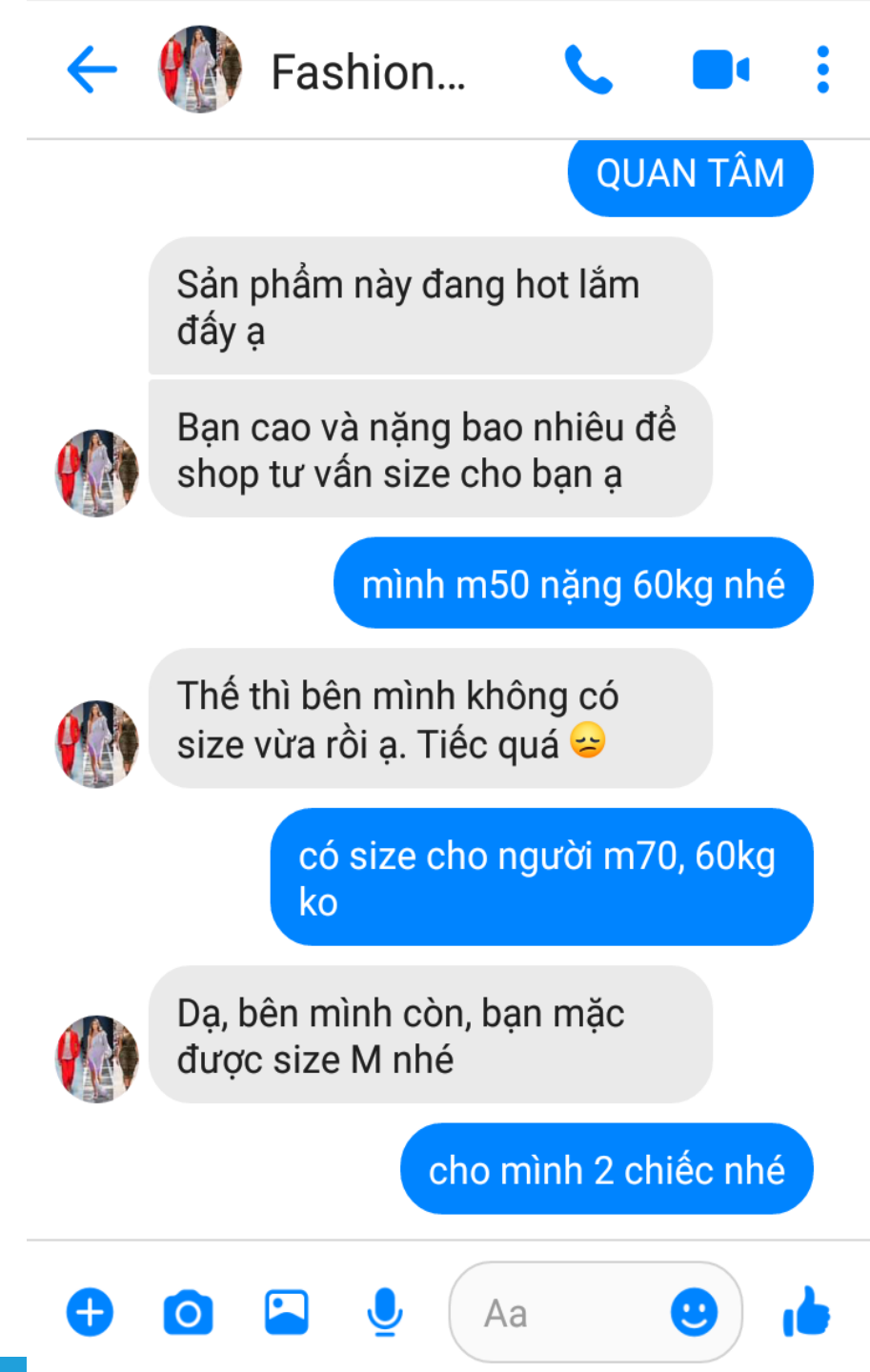
Giới thiệu về Chatbot





Chatbot là gì?

Là một chương trình máy tính tương tác với người dùng dưới một giao diện như âm thanh hoặc tin nhắn.





Nguyên nhân hình thành

53 triệu Tài khoản/100 triệu dân Việt Nam

3h/ngày Online facebook trung bình của người dùng Việt Nam

70% Các thắc mắc, vấn đề của người dùng bị trùng lặp

40% khách hàng mong muốn được tư vấn trong 1h trở lại

90% Người dùng mong đợi doanh nghiệp cung cấp dịch vụ tự chăm sóc khách hàng



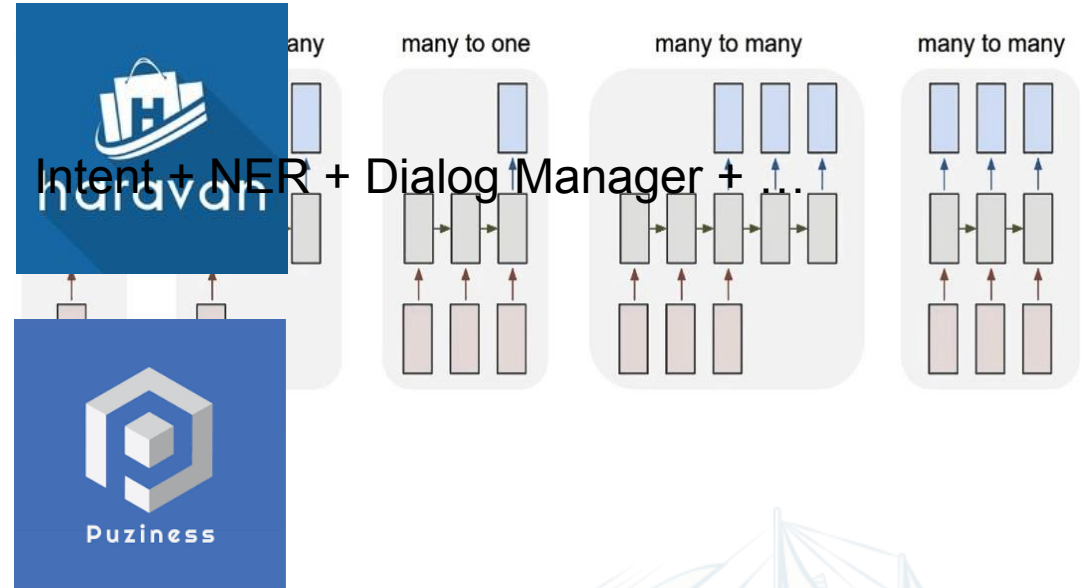


Phân loại Chatbot

- Chatbot hướng Generative
- Chatbot hướng Kịch bản
- Chatbot hướng Ngữ cảnh



Kiến trúc seq2seq
RNN / LSTM + Attention





Case Study thực tế





Chatbot cho thời trang

- Chatbot tư vấn cho khách hàng về các chính sách về shop
- Chatbot tư vấn cho khách hàng về kích thước, màu sắc, giá cả, thương hiệu ...
- Chatbot tư vấn size cho khách hàng
- Chatbot tự động chốt đơn hàng



Các khó khăn và hướng giải quyết

- Dữ liệu => Bắt buộc làm thủ công
- Nghiệp vụ
- Lượng request lớn => Microservice hoá các service, tăng tính mở rộng cao
- Độ tin cậy cao => Cải thiện mô hình dựa vào các nhân viên bán hàng





Kiến trúc tổng quan





Giải pháp công nghệ

Programming Languages



Framework + Lib



gunicorn



Databases



Others



logstash



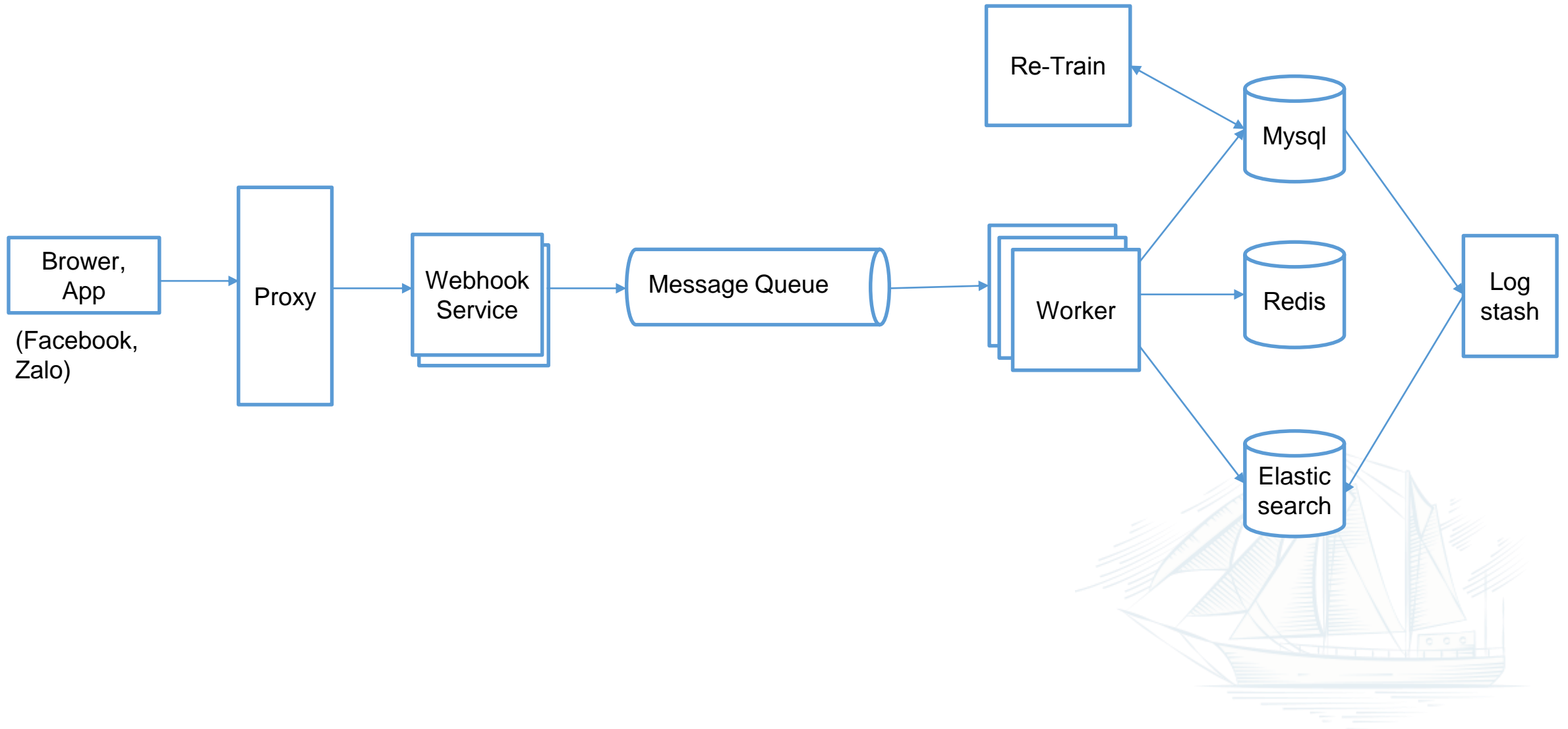
elasticsearch

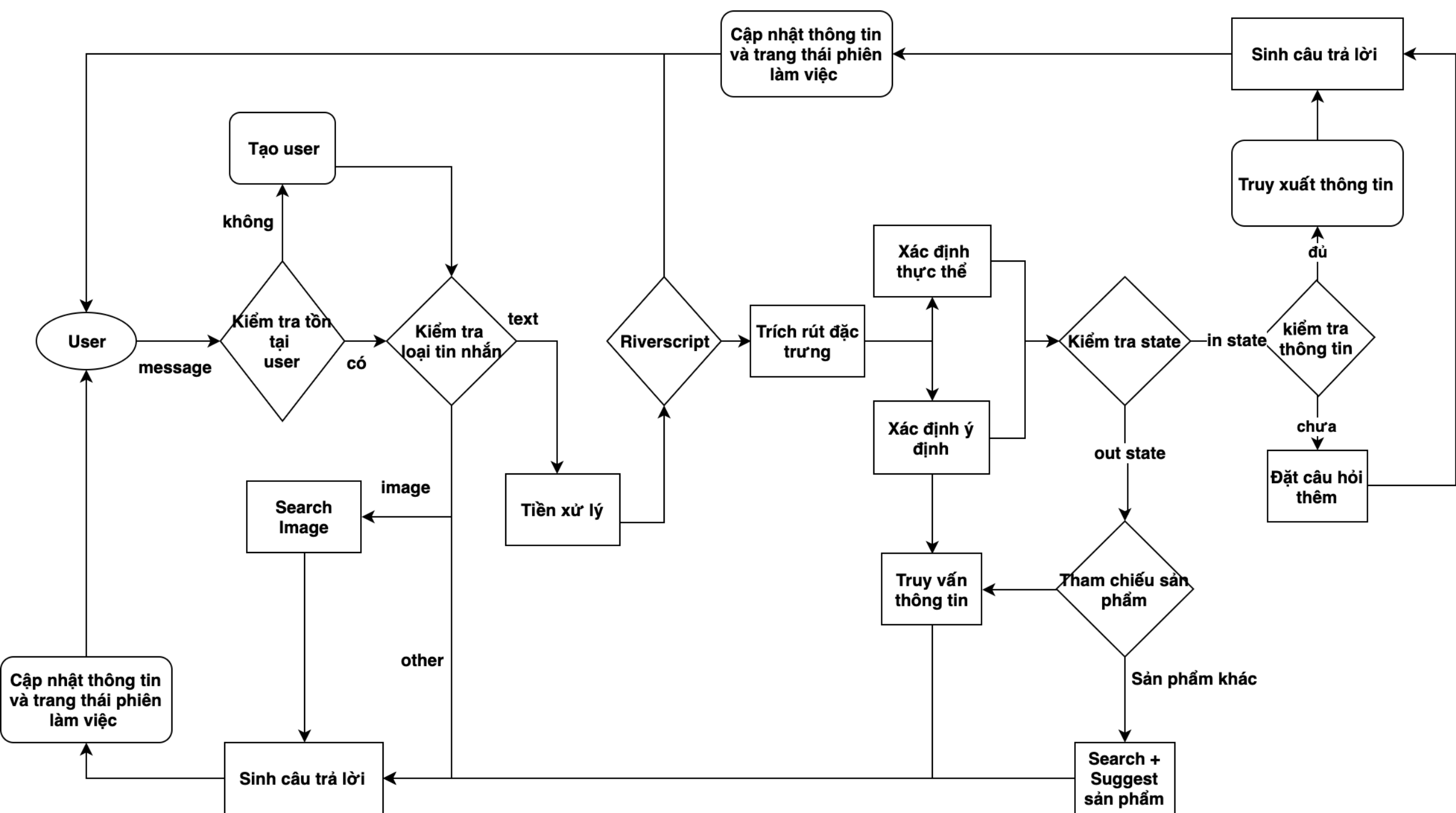
Web Servers



Message Queues









Preprocessing

Mức 1:

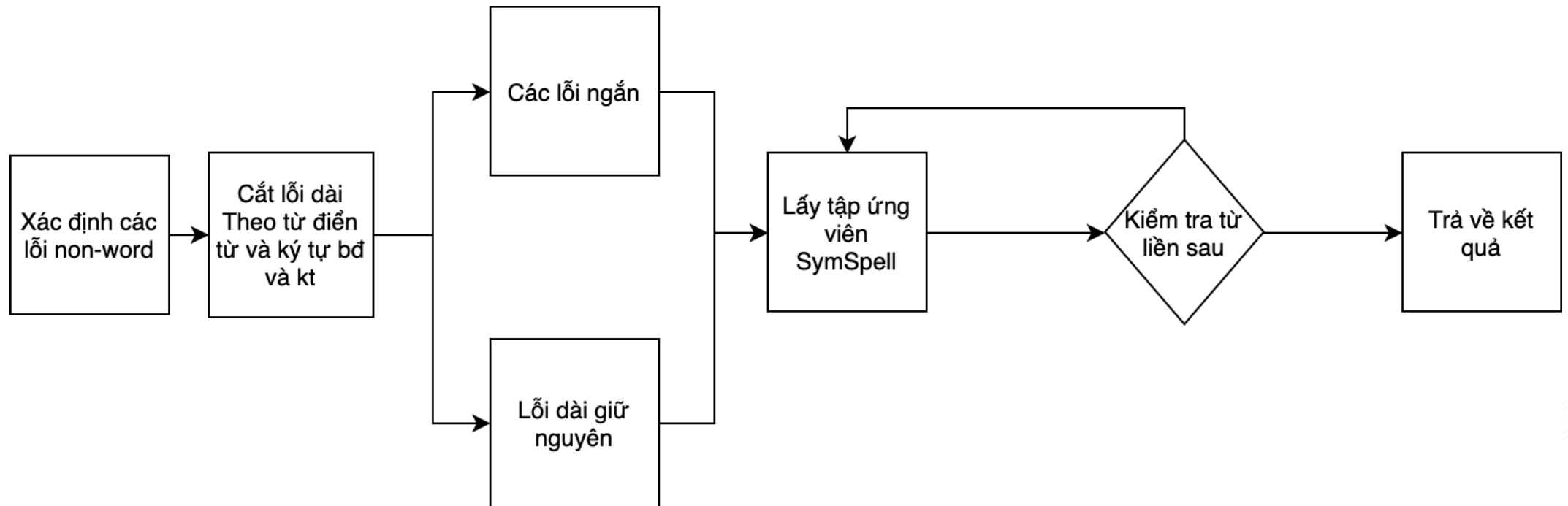
- Chuẩn hoá unicode (tổ hợp => dụng sẵn)
- Chuẩn hoá vị trí dấu: òa -> oà
- Thay thế từ viết tắt: sử dụng từ điển
- Xoá các icon, ký tự đặc biệt: 🙄❤️,!,#...





Preprocessing

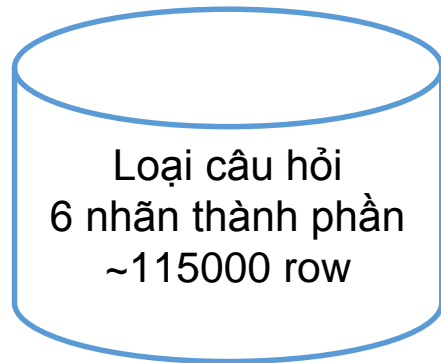
Mức 2:





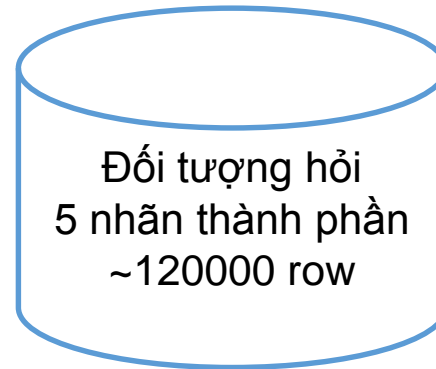
Intent

- Mục đích: Xác định được ý định của người dùng để chatbot có thể đưa ra hỗ trợ chính xác
- Dữ liệu: 5000 Cuộc hội thoại



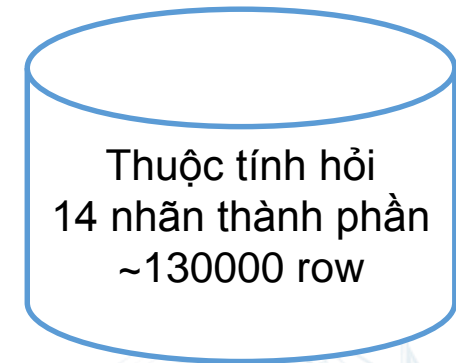
Xác định câu trả lời
phù hợp, kiểu truy vấn

Ex: Yes/no, attribute,
exists...



Xác định đối tượng
truy vấn

Ex: Shop, Product, Ship,
Promotion ...



Thuộc tính
truy vấn

Ex: Price, Color, Size ...



Intent

Feature: BOW (Bag Of Word) + TF-IDF + Bi-Gram

Một số kết quả đối sánh giữa các thuật toán

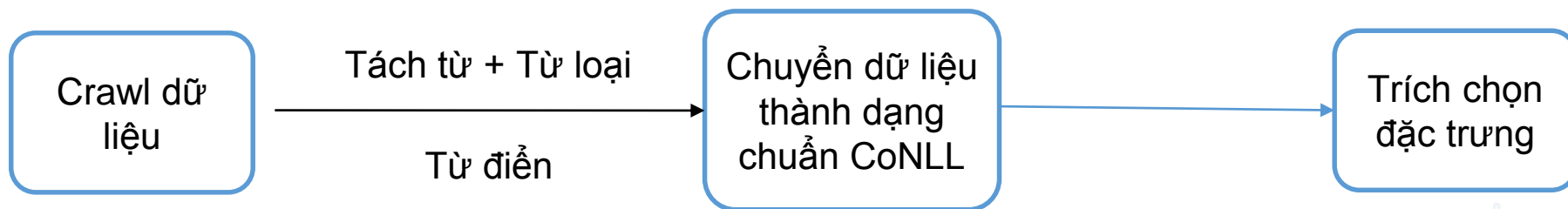
Thuật Toán \ Đánh giá	Đánh giá		
	Avg-Precision	Avg-Recall	Avg-F1
LogisticRegression	0.89	0.88	0.88
SGDClassifier	0.90	0.89	0.90
AdaBoostClassifier + SVC	0.92	0.93	0.93



Named Entity Recognition

- Mục đích: Xác định thực thể trong câu
- Dữ liệu: 5000 Cuộc hội thoại

18 Tag: *COLOR, SIZE, PRICE, GENDER, LOCATION, TIME, WEIGHT, HEIGHT, QUANTITY ...*



shop	N	B-NP	O	O
có	V	B-VP	O	O
áo	N	B-NP	B-TYPE	O
kẻ	V	B-VP	O	O
đen	A	B-AP	B-COLOR	O
trắng	A	B-AP	B-COLOR	O
không	R	O	O	O
ạ	T	O	O	O





Named Entity Recognition

Loại	Đặc trưng	Giải thích
Mẫu ngữ cảnh cơ bản	$[-2,-1,0,1,2]$ $[(-2,-1), (-1,0), (0,1),(1,2)]$	Duyệt 5 từ hoặc cụm 2 từ liên tiếp
	lower, istitle, iscontainnumber	Kiểm tra thuộc tính của từ hiện tại: Lấy giá trị viết thường của từ, Viết hoa chữ cái đầu, Chứa số
Mẫu ngữ cảnh từ điển	is_in_dict, is_in_dict_loc	Kiểm tra xem từ hiện tại có thuộc trong từ điển tiếng Việt, từ điển vị trí
Mẫu ngữ cảnh từ loại	$[-2,-1,0,1,2][1]$ $[(-2,-1), (-1,0), (0,1),(1,2)][1]$	Lấy giá trị từ loại của từ theo ugram và bigram



Named Entity Recognition

Kết quả training với 500 epochs

Đánh giá Thuật Toán	Avg-Precision	Avg-Recall	Avg-F1
CRF	0.96	0.97	0.96

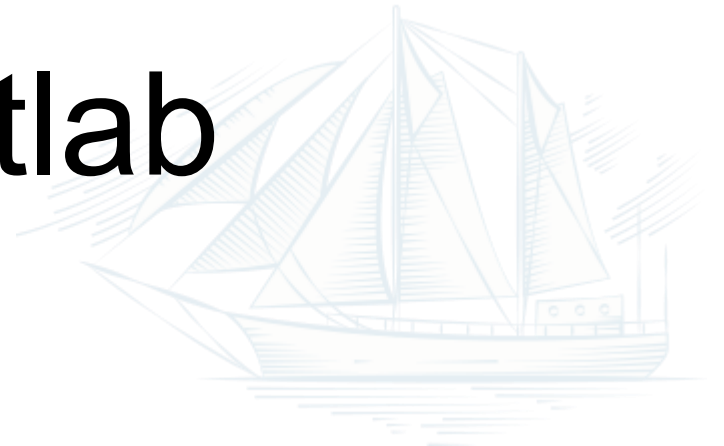
Kết hợp with regex:

- Weight: `(\d+(?:[.,]\d+)?[]*(?:[kK][gG]))`
- Height: `(\d+(?:[.,]\d+)?[]*(?:[cC]?[mM])\d*)`





CI/CD với Docker & Gitlab





Dockerfile cho mọi service

```
1 FROM python:3.7-alpine
2
3 # Set the working directory to /app
4 WORKDIR /app
5
6 COPY requirements.txt /app
7
8 # Install any needed packages specified in requirements.txt
9 RUN pip install --no-cache-dir -r requirements.txt
10
11 # Copy app source code
12 COPY . /app
13
14 # Alias log to output
15 RUN mkdir -p /app/logs \
16     && ln -nfs /dev/stdout /app/logs/info.log \
17     && ln -nfs /dev/stderr /app/logs/error.log
18
19 EXPOSE 5000
20
21 CMD gunicorn -b 0.0.0.0:5000 -w 2 wsgi:app --log-file -
```





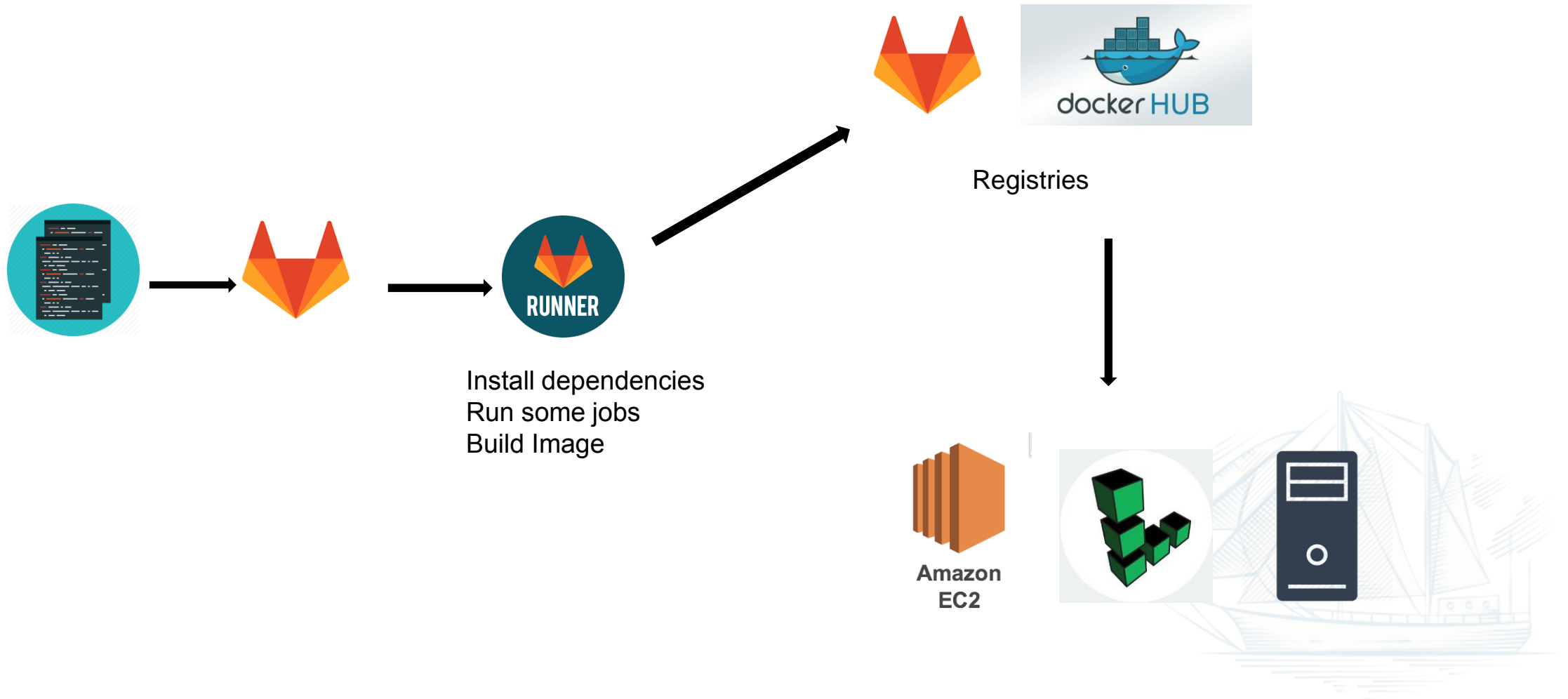
CI/CD tự động với Gitlab

- Setup GitlabRunner
- .gitlab-ci.yml

```
38  #=====
39  # Define jobs
40  #=====
41  coding-style:
42  |   extends: .test
43  |   script:
44  |     - pylint src --output-format=colorized --disable=duplicate-code,too-few-public-methods,no-self-use --ignore test
45  |
46  unit-test:
47  |   extends: .test
48  |   script:
49  |     - echo "TODO run unit-test"
50
51  latest-image:
52  |   stage: build
53  |   script:
54  |     - docker pull $CI_REGISTRY_IMAGE:latest || true
55  |     - docker build --pull --cache-from $CI_REGISTRY_IMAGE:latest -t $CI_REGISTRY_IMAGE:latest .
56  |     - docker push $CI_REGISTRY_IMAGE:latest
57  |   only:
58  |     - master
```



CI/CD tự động với Gitlab





Các công cụ hỗ trợ







Các sản phẩm khác





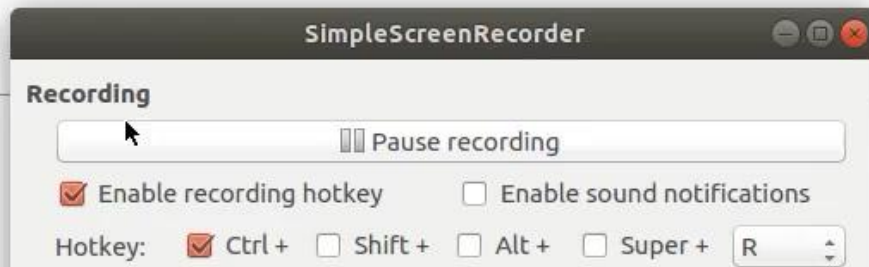
Spell Check

Spell Checker - Rabiloo Inc

Trước sự chugnws kiến của ông Nguyễn Thế Hoàn, Bí thư Huyện ủy Lộc Hà và chính quyền địa phương, Tiến sĩ Đoàn Hoài Sơn mang đến cho em Hiền và gia đình một thông tin vui. “Hôm nay chúng tôi có mặt ở đây, trước nữa là để động viên sẽ chia với mất mát, đau thương khi cháumất đingười bố, rồi còn phải cahwm mẹ bệnh nặng và 2 em còn nhỏ; sau đó là thông tin quyết định tiếp nhận cháu vào theo học ở trường” - thầy Sơn thông tin tới gian đình em Hiền. Lời gợi mở của lãnh đạo Trường ĐH Hà Tĩnh nhận được nhiều ý kiến đồng tình của lãnh đạo huyeenjxã và nguiowfthân của em Hiền. “Chú nghĩ rằng lời gợi mở của thầy Sơn là rất hợp lí, cháu nên suy nghĩ về phương án này. Học tại Hà Tĩnh, không chỉ gần chăm sóc mẹ và các em, cháu còn có cơ hội tìm được việc làm sau khi ra trường” - ông Hoàn đưa ra lời góp ý với Hiền. Cùng với Trường ĐH Hà Tĩnh, đã có nhiều truognwfdaij học khác trên cả nuocwis (trong đó có ĐH QG HN, ĐH QG TP. HCM) đã liên hệ với Hiền thông báo sẵn sàng tiếp nhậm em. Có lúc Hiền đã phải xin nghỉ học để chăm bố mẹ và các em, nhưng tiếc cho một em học sinh nghèo học giỏi vượt khó nhiều năm, thầy cô giáo và các bạn học sinh trường THPT Nguyễn Văn Trỗi đã giúp đỡ hỗ trợ em tieps tục đến trường với bao ucows mơ hoài bão trở thành cử nhân kinh tế tuongwlai. Nhưng số phận thật nghiệt ngã, đúng vào ngày thi đầu tiên của Kỳ thi tốt nghiệp THPT 2019, bố Hiền đột ngột qua đời khiến em phải bỏ thi để chịu tang bố.

Check

Done





Detect Location

Postman interface showing a REST client setup for a "detect_location" endpoint.

Endpoint: POST `{{host}}/api/location`

Request Body (JSON):

```
1 {
2   "text": ""
3 }
```

Response (JSON):

```
1 {
2   "addresses": [
3     {
4       "address": "7 Vô Công Tôn",
5       "case": "steward_districtprovince_",
6       "confident": 1.0,
7       "district": "Tân Phú",
8       "province": "Hồ Chí Minh",
9       "ward": "Tân Quý"
10    },
11    {
12      "address": "7 Vo Cong Ton Tan Quy",
13      "case": "districtprovince_",
14      "confident": 1.0,
15      "district": "Tân Phú",
16      "province": "Hồ Chí Minh",
17      "ward": ""
18    }
19  ]
20 }
```

Status: 200 OK | **Time:** 142ms | **Size:** 679 B | **Save Response**



Search and Detect Image

Chrome File Edit View History Bookmarks People Window Help

img_5134_copy_cop x AI Rabiloo x Váy xòe xỏp cộc tay x New Tab x AI Rabiloo - Screenc x AI Rabiloo.webm - G x AI Rabiloo - Screenc x

https://48d8ed05.ngrok.io/image-search

AI RABILOO

- Spell Checking
- Entity Type
- Intent
- Chatbot
- Image Search**
- Image Detect

Image Search

Demo Search Image for Shops of VietNam

Drag & Drop files here

or

Link image

Browse files

SEARCH

powered by Screencastify

HOME COMPANY PORTFOLIO BLOG

© 2019 Product of Rabiloo

img_5134_copy....jpg

Screencastify - Screen Video Recorder is sharing your screen. Stop sharing Hide

Show All x



Thank you!

Q&A

