

# HƯỚNG DẪN GÁN NHÃN DỮ LIỆU NLP

## ***Giới thiệu:***

Gắn nhãn dữ liệu với mục đích phát triển mô hình Xử lý ngôn ngữ tự nhiên cho ứng dụng Chatbot.

## ***Các loại nhãn:***

### **Mô tả:**

Sentence label: Nhãn cho cả một item.

Entity label: Nhãn cho mỗi cụm từ trong câu.

### **Định nghĩa chi tiết:**

#### Intent (Sentence Label):

- **symptom**: câu hỏi về dấu hiệu/triệu chứng của bệnh( Cho mình hỏi dấu hiệu,triệu chứng của bệnh ung thư não là gì)
- **overview**: Câu hỏi tổng quát về bệnh( bác sĩ cho tôi hỏi bệnh đột quỵ là gì)
- **method diagnosis**: Câu hỏi về các phương pháp chẩn đoán bệnh ( Cho tôi hỏi chẩn đoán bệnh ung thư gồm những gì)
- **severity**: Câu hỏi về mức độ của bệnh ( bác sĩ cho tôi hỏi bệnh đột quỵ có nguy hiểm không)
- **treatment**: Câu hỏi về điều trị bệnh ( Cho tôi hỏi việc điều trị ung thư gồm những gì)
- **cause**: Câu hỏi về nguyên nhân gây ra bệnh ( Bệnh ung thư phổi gây ra bởi những nguyên nhân nào)

#### Named Entity:

- **disease**: Tên các bệnh( Ví dụ Ung thư trực tràng, Bệnh tim mạch,Covid...)
- **symptoms**: Các dấu hiệu của bệnh ( đau đầu, sốt cao, nhức mỏi ...)
- **medical value**: Các thông số, giá trị ( huyết áp 120-129 mmHg, nhịp tim 90 nhịp/phút...)
- **severity**: Mức độ của bệnh( Nguy cấp, nghiêm trọng, bình thường, giai đoạn 1, giai đoạn 2...)
- **medical procedure**: Thủ tục y tế( xét nghiệm, chụp x quang)
- **individual behavior**: Hành vi ( uống thuốc x, ăn phải nấm độc...)
- **location**: địa điểm (Hà Nội, quận x...)
- **ORG**: Tổ chức(Bệnh viện vinmec,bệnh viện y dược...)
- **time**: Thời gian (11h sáng, 3h chiều...)
- **date**: ngày tháng năm( 11/3/2021 ...)
- **gender**: Giới tính (nam nữ..)
- **age**: Tuổi
- **person name**: Tên người

#### Relation:

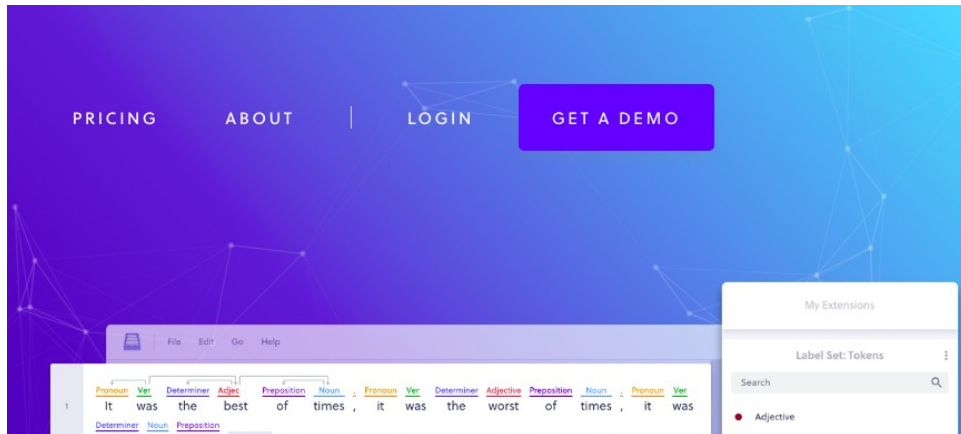
- **has**
- **when**
- **where**
- **of**

## Hướng dẫn thực hiện:

### 1. Đăng nhập

Truy cập liên kết: <https://datasaur.ai/>

#### 1. Click vào Login ở góc trái trên màn hình



Hình 1: Giao diện trang chủ công cụ gán nhãn dữ liệu Datasaur

### 2. Điền username và password để đăng nhập

A screenshot of the Datasaur.ai login page. At the top is the Datasaur.ai logo. Below it are two input fields: "Email \*" with the value "v.huyta@vinbrain.net" and "Password \*" with a masked password "\*\*\*\*\*". Below the password field is a link "Forgot password? Reset". Below that is a reCAPTCHA widget showing a green checkmark and the text "I'm not a robot". At the bottom is a large blue "Sign In" button. Below the button is a link "or sign in with" followed by a social media login button.

Hình 2: Giao diện đăng nhập

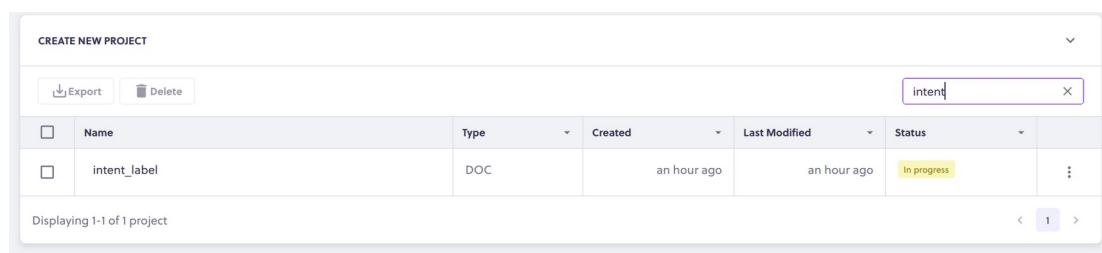
## 2. Chọn Project

1. Nhấp vào khung tìm kiếm bên phải màn hình và nhập vào tên dự án được phân công để tìm kiếm



Hình 3: Giao diện tìm kiếm dự án

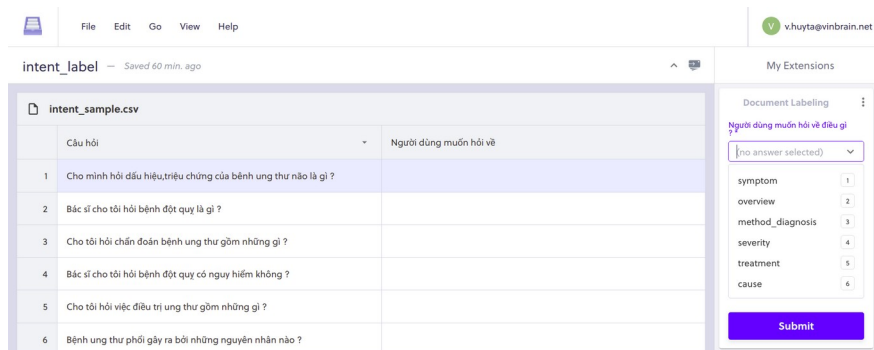
## 2. Chọn project mà mình được phân công, lưu ý chọn đúng Type



Hình 4: Giao diện chọn dự án thực hiện, ví dụ intent\_label là tên dự án mẫu

## 3. Label cho Intent

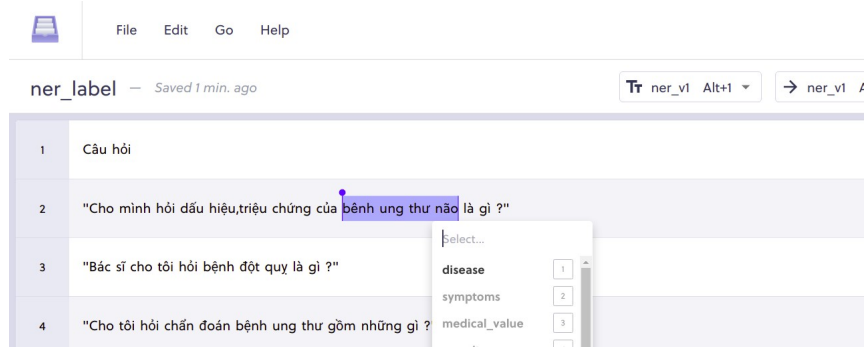
1. Với mỗi câu, click vào hàng cho câu hỏi. Chọn kết quả tương ứng trong danh sách phía bên tay phải màn hình. Để kết thúc việc gán nhãn cho hàng dữ liệu hiện tại bấm “Submit”.



Hình 5: Giao diện gán nhãn dữ liệu theo câu

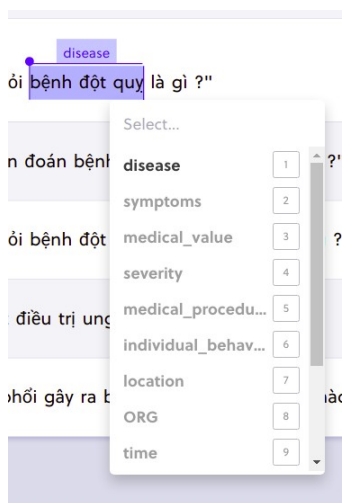
## 4. Label cho Named Entity và Relation.

1. Với mỗi câu, kéo chuột để chọn cụm từ tương ứng cần gán nhãn.

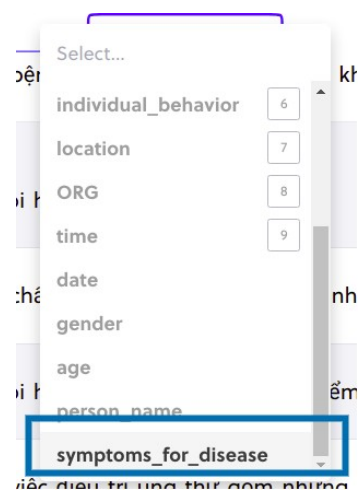


Hình 6: Giao diện gán nhãn dữ liệu theo từ/cụm từ

2. Chọn nhãn tương ứng cho cụm từ trong danh sách trả về.

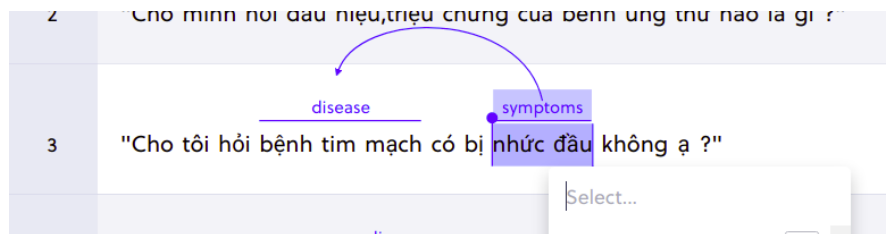


Hình 7: Danh sách nhãn cho các từ cần gán nhãn



Hình 8: Giao diện lựa chọn nhãn cho mối liên hệ giữa từ/cụm từ với nhau

3. Kéo chuột từ cụm từ này đến cụm từ khác **theo đúng chiều**. Chọn mối quan hệ tương ứng giữa 2 cụm từ.



Hình 9: Giao diện kéo thả chuột biểu diễn mối quan hệ giữa các từ/cụm từ