



# Exploring Diseases Dataset for Chatbot Dialogue Management

---

Tai Ng. (Applied Scientist)

April 15, 2021

VINBRAIN INTERNSHIP PROGRAM 2021

## Task-Oriented Dialogue Systems

---

- Help users achieve their specific goals.
- User forgets some information.
- Focus on understanding users, tracking states, and generating next actions.
- Minimize the number of turns: fewer turns the better.

## Diseases Dataset

A dataset to provide the students a source to create a healthcare related system.

	Disease	Symptom_1	Symptom_2	Symptom_3	Symptom_4	Symptom_5
0	Fungal infection	itching	skin_rash	nodal_skin_eruptions	dischromic_patches	NaN
1	Fungal infection	skin_rash	nodal_skin_eruptions	dischromic_patches	NaN	NaN
2	Fungal infection	itching	nodal_skin_eruptions	dischromic_patches	NaN	NaN
3	Fungal infection	itching	skin_rash	dischromic_patches	NaN	NaN
4	Fungal infection	itching	skin_rash	nodal_skin_eruptions	NaN	NaN
...	...	...	...	...	...	...

Figure 1: Diseases Dataset

## Insight

---

- Dataset has 4920 rows, but it has 41 distinct diseases.
- Synthesizing from multiple sources, symptoms are overlap.

→ Re-architecting dataset

## Database

---

```
/* 1 */
{
  "_id" : ObjectId("606c0a179dad241387f678c"),
  "Disease" : "(vertigo) Paroxysmal Positional Vertigo",
  "Symptom" : [
    "spinning_movements",
    "unsteadiness",
    "loss_of_balance", |
    "vomiting",
    "headache",
    "nausea"
  ]
}

/* 2 */
{
  "_id" : ObjectId("606c0a179dad241387f678d"),
  "Disease" : "Alcoholic hepatitis",
  "Symptom" : [
    "distention_of_abdomen",
    "history_of_alcohol_consumption",
  ]
}
```

**Figure 2:** Disease database

## Heuristic

---

Heuristic Search Method

## Agent's action

---

1. Inform & Request: Amount of records greater than one.
2. Match found: Only one record satisfy user's constraints.
3. Done: End conversation if cannot found disease.

## Demo

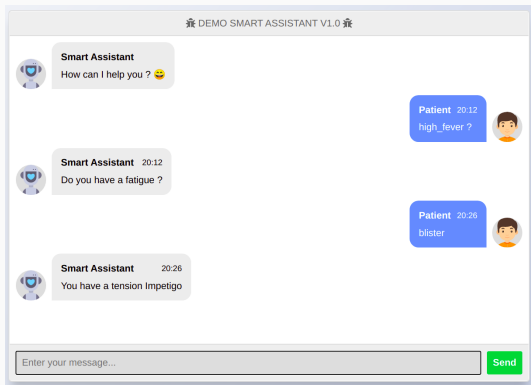


Figure 3: Demo conversation



Q&A