



# 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.

## Conversation

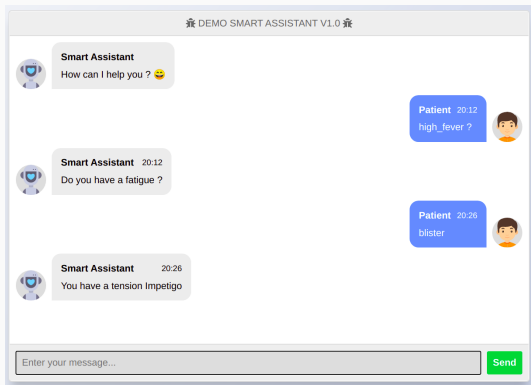


Figure 3: Demo conversation



## User's Semantic Frame

---

```
=====
USER'S ACTION: {'intent': 'request', 'inform_slots': {'Symptom': ['high_fever']}, 'request_slots': {'Disease': 'unk'}}
=====
```

**Figure 4:** Generate semantic frame from user's action

## Ranking Symptoms

```
=====
DICTIONARY VOTE SYMPTOM: {'high fever': 12, 'fatigue': 9, 'vomiting': 6, 'chills': 6, 'headache': 5, 'malaise': 5, 'nausea': 4, 'cough': 4, 'loss of appetite': 4, 'swelled lymph nodes': 3, 'breathlessness': 3, 'muscle pain': 3, 'phlegm': 3, 'skin rash': 3, 'chest pain': 3, 'abdominal pain': 3, 'sweating': 3, 'yellowing of eyes': 2, 'joint pain': 2, 'itching': 2, 'yellowish skin': 2, 'red spots over body': 2, 'mild fever': 2, 'weight loss': 2, 'diarrhoea': 2, 'dark urine': 2, 'red sore around nose': 1, 'acute liver failure': 1, 'toxic look (typhos)': 1, 'family history': 1, 'redness of eyes': 1, 'yellow crust ooze': 1, 'stomach bleeding': 1, 'coma': 1, 'loss of smell': 1, 'constipation': 1, 'lethargy': 1, 'extra marital contacts': 1, 'congestion': 1, 'pain behind the eyes': 1, 'throat irritation': 1, 'mucoid sputum': 1, 'belly pain': 1, 'blister': 1, 'back pain': 1, 'blood in sputum': 1, 'fast heart rate': 1, 'continuous sneezing': 1, 'runny nose': 1, 'patches in throat': 1, 'sinus pressure': 1, 'rusty sputum': 1, 'muscle wasting': 1}
=====
```

Figure 5: Count frequency of symptoms in query diseases

Q&A