

USER GENERATED CONTENT (UGC)

Introduction :

UGC, user generated content, as evident from the name is generated by the user when he/she visits a medicine page. It consists of series of questions in the various formats such as single-select, multi-select, free text and slider . Our main objective was to convert the free text data format into multi- select auto suggest format for the following reason:

- To make data more structured.
- Convenient for the user.
- Less noise overall giving quality text data.

Free text answer corresponds to three different questions that are -

1. Any specific experiences or tips regarding side-effects of this medicine?
2. Are there are any other specific tips or experiences related to \$name that you would like to share?'
3. Any specific tips or dietary precautions with this medicine?

Since first two questions are almost same, we divided the free text data into two categories i.e questions related to side effects and that related to the dietary precautions the user has taken.

UGC of Side Effects :

In order to do so, a dictionary is made which maps the entity_id i.e the product/medicine to the most common side effects faced by the users.

```

66181: ['obesity'],
66601: [('headache', 'fever'), 'headache', 'fever'],
66632: ['bloating', 'weakness'],
67303: ['bleeding'],
67304: ['sweating', 'pain'],
67577: ['nausea'],
68437: [],
68483: [('sugar', 'level')],
68677: [],
69349: [],
69489: [],
69771: [('throat', 'infection'), 'infection'],
70139: ['cramps'],
70548: ['rash'],
70696: [],
70825: ['constipation'],
70901: [],
71063: ['ache'],

```

We also encountered some missing values due to the lack of data for a particular entity id which can later be auto filled.

Also, the text data of UGC can be mapped in a following way to the side effects corresponding to it :

```

'dear sir experience severe cramps legs eat medicine': ['cramps']
'drug causes swelling feet': ['swelling']
'loss appetite dry mouth etc': [('loss', 'appetite'), ('dry', 'mouth')]

```

Examples of tags related to 'Side Effects' are as follows :

```

['depression']
[('abdominal', 'pain'), 'pain', 'diarrhoea']
['cough', 'vomiting']
['pain']
['gastritis']
[('skin', 'allergy'), 'allergy', 'cough']
['vomiting']
['trembling']
['drowsiness']
['dizziness']

```

['itching']
['stroke']

UGC of Dietary Precautions :

Similarly, a dictionary that maps the entity id to the dietary precautions that should be taken by the users :

```
66632: ['low', 'fat', 'less', 'spicy'],  
67307: [],  
67577: ['oil'],  
70206: [],  
70696: [],  
71063: ['avoid', 'alcohol', 'spicy'],  
74710: [],  
110605: [],  
113708: ['non', 'veg'],  
113791: ['normal'],  
114065: [],  
116492: [],  
117079: [],  
117234: ['low', 'fat'],  
117239: [],  
118330: ['milk'],
```

Text data can be mapped as -

```
'avoid alcohol spicy foods': ['avoid', 'alcohol', 'spicy'],  
'avoid milk products doctor advice': ['avoid', 'milk'],  
'avoided much oil ghee': ['avoided', 'oil'],
```

The problem that we encountered while dealing with UGC related to dietary precautions are as follows -

1. Data is not enough to be a reliable. Only 40 UGC's gave the consequential meanings.
2. Sometimes we were unable to capture the negation of the statements or to capture the complete meaning of it. Example of such cases are -

'sugar free diet ': ['sugar']

'ya salt restricted' : ['salt']

Tags related to diet are as follows -

['regular']
['oil']
['salt']
['empty', 'stomach', 'alcohol']
['sugar']
['milk']
['alcohol']
['spicy']
['oily', 'fatty', 'non', 'veg']
['alcohol', 'spicy']
['normal']
['fat', 'spicy']
['fat']
['non', 'veg']
['fatty', 'oily']
['veg']
['normal', 'veg']

Suggestions -

- UGC of Side Effects :

1. ' No Side Effects' option should be available for every product.
2. The side effects that are available corresponding to each entity id should be included in the multi-select feature.
3. An option of ' None of the above' which will enable the user to write the free text description. This will eventually help us in getting more data which can lead to more robust multi-select options in the future.

- UGC of Dietary Precautions :

1. ' Normal Diet' option should be there for every product.
2. Only reliable dietary precautions can be added in multi-select format corresponding to the entity id.

Alternative is to ask these two questions related to dietary precautions and side effects for every product. In dietary precautions, there will be two options -

1. Normal Diet
2. An autofill search space for user such that it displays " Should be taken with ____ ". The blank space maps to the 'diet' tags available from the UGC data.
3. An autofill search space for user such that it displays " Avoid ____ ". Blank space will be filled in a similar way.

Similarly in case of side effects -

1. No side effects
2. Autofill search space which maps to the 'side effects' tags available from the UGC data.