

## Q1: Query and Document Matching

It is very common in Spoonshot to map food ingredients to articles, and find which ingredients are most relevant in a particular article.

This specific problem has come out of a generic class of problems called “Search” where several search algorithms are built to match relevant documents to a given query. You can think of what Google Search or Bing search does.

Here, as part of this problem you will be provided a list of ingredients and an article. You need to return a ranked list of relevant ingredients for this given article. You are encouraged to augment the list of ingredients with more open-source information like description etc.

I tried solving this problem using different methods. I will mention a few below:

### Method 1

- A score of 1 will be given if we are able to find an exact match of a query in the article.
- I have also made sure to give partial scores in case of partial matching. From this I was able to generate a ranking as shown below:(Left side is the query and on right side is their score)

```
{'Tomato': 1.0,  
'exotic fruit': 1.0,  
'tea': 1.0,  
'Ambrette Seed': 0.5,  
'Baby Abalone': 0.5,  
'Campari Tomato': 0.5,  
'Chia Meal': 0.5,  
'Munch Chocolate': 0.5,  
'Organic Maize': 0.5,  
'Vegan Chorizo': 0.5,  
'Vegan Marshmallow': 0.5,  
'Vegetable Stock': 0.5,  
'Organic Peruvian Groundcherry': 0.3333333333333333,  
'Organic Tartar Cream': 0.3333333333333333,  
'Vegan Carob Chips': 0.3333333333333333,  
'Vegan Chicken Strips': 0.3333333333333333,  
'Vegan White Cake': 0.3333333333333333,  
'Vegan Puff Pastry Sheet': 0.25,  
'Vegan Semisweet Chocolate Chips': 0.25}
```

### Method 2

- A score of 1 will be given if we are able to find a similarity match(based on get\_close\_match module from difflib library).
- Multiple occurrences can boost the score again by 1.Using this the following ranking was obtained.

```
{'Vegan Chorizo': 4.5,
'Vegan Marshmallow': 4.5,
'Vegan White Cake': 3.666666666666667,
'Chia Meal': 3.0,
'Vegan Carob Chips': 3.0,
'Vegan Chicken Strips': 3.0,
'Vegan Puff Pastry Sheet': 2.5,
'Vegan Semisweet Chocolate Chips': 2.5,
'Munch Chocolate': 2.0,
'Celery Soup': 1.5,
'exotic fruit': 1.5,
'Snickers Spread': 1.5,
'Crunch Bars': 1.0,
'Organic Tartar Cream': 1.0,
'Pork Tripe': 1.0,
'Organic Maize': 1.0,
'tea': 1.0,
'Tomato': 1.0,
'Vinegar': 1.0,
'Apple Cinnamon Granola': 0.6666666666666666,
'Organic Peruvian Groundcherry': 0.6666666666666666,
'Baby Abalone': 0.5,
'Strawberry Gelatin': 0.5,
'Ambrette Seed': 0.5,
'Pork Lungs': 0.5,
'Campari Tomato': 0.5,
'Pork Chump Chops': 0.3333333333333333}
```

### Method 3

- This is implemented using a transformer based model from huggingface. The model uses zero-shot learning. Using this pretrained model I was able to generate a ranking as shown below. It shows the top 20 classes.

```
'labels': ['Organic Maize',
'Murukku Packet',
'Vegetable Stock',
'Giardiniera',
'Vinegar',
'Mango',
'Smokies Sausage',
'Muesli',
'Tamarind',
'Cardamom',
'Ambrette Seed',
'Snickers Spread',
'Mulberry',
'Hog Maw',
'Chia Meal',
'Organic Peruvian Groundcherry',
'Celery Soup',
'Orange Extract',
'Vegan Puff Pastry Sheet',
'Vegan Chorizo',
'Vegan Marshmallow',
'Pickled Cauliflower',
```

## Conclusions:

- The first 2 methods is based only on the semantic meaning and seems to perform okay in the given case.
- The transformer based model will perform better generally as it considers both sentimental and semantic representations of words.

### Q2: How you code and solve problems

Given there is an array, write a function that returns as follows -

Input : array = [5, 1, 4, 2]  
Output : [8, 40, 10, 20]

Input : array = [1, 0, 3, 4]  
Output : [0, 12, 0, 0]

You will be evaluated on -

- Finding the pattern between the input and the output.
- Writing the executable code.
- Code standards

- Given an integer array of nums, we want to output an array answer such that  $\text{res}[i]$  is equal to the product of all the elements of nums except  $\text{nums}[i]$ .
- Code snippet

```
1  def productExceptSelf(nums):
2      res=[1]*len(nums)
3
4      pre=1
5      for i in range(1,len(nums)):
6          res[i]=pre*nums[i-1]
7          pre=res[i]
8      post=nums[-1]
9
10     for i in range(len(nums)-2,-1,-1):
11         res[i]=post*res[i]
12         post=post*nums[i]
13
14     return res
15
16     n = int(input("Enter number of elements : "))
17     nums=[]
18     for i in range(0, n):
19         ele = int(input())
20         nums.append(ele)
21
22     print(productExceptSelf(nums))
```

### Q3: What's your motivation ?

- As an intern in Spoonshot what do you want to achieve in 8 weeks ?
- What is your motivation in the field of Data Science / Machine Learning ?
- If you are given a full time offer as a Data Scientist in Spoonshot what is your dream problem which you would like to solve ?

- I want to gain Industrial experience in Data Science. I am really looking forward to learning new skills and gaining knowledge. I want to be part of a complex ML/DS Project that can challenge me. I want to interact and learn from the top data scientists at Spoonshot which will help me in my journey toward becoming a Data Scientist.
- I am a person who likes to be productive, innovative, and results oriented. From the Projects that I completed, I feel that data Science is a field that enables me to be really creative. I know that Data Science when used correctly can solve any complex problem known to us. Statistics and Probability are my favourite topics in mathematics which are extensively used in Machine Learning. I also like programming. All these interest aligned

together to make me do projects and courses in Data Science, which further enhanced my excitement about it.

- c. I would like to problems related to predicting future trends in food culture. This looks like an interesting and complex problem that can help me learn a lot. Given a chance, I would also like to solve NLP problems.