**Submission 1 Explaination Document**

*Team AIvengers*

Akshay Kumar

Anubroto Ghose

Devina Goel

Our approach for: -

item\_volume,

item\_weight,

maximum\_weight\_recommendation,

voltage and

wattage

was the same. We noticed that a simple text recognition (OCR) would be enough in most of the cases to extract the valuable information. We have the entity\_name which is what is used to understand what kind of keywords we are searching for in the OCR output text  
  
After getting the OCR text, we apply NLP and RegEx techniques to extract the necessary details when present.

For Item\_weight - unit\_map = {

'gram': ['g', 'gram', 'gm'],

'kilogram': ['kg', 'kilogram'],

'microgram': ['mcg', 'microgram', 'μg'],

'milligram': ['mg', 'milligram'],

'ounce': ['oz', 'ounce'],

'pound': ['ibs', 'Ibs', 'lb', 'lbs', 'pound'],

'ton': ['ton', 't']

}  
We are looking for these keywords in the OCR text. When found we extract the value and the unit and return the answer

For maximum\_weight\_recommendation- unit\_map = {

'gram': ['g', 'gram', 'gm'],

'kilogram': ['kg', 'kilogram', 'kcs'],

'microgram': ['mcg', 'microgram', 'μg'],

'milligram': ['mg', 'milligram'],

'ounce': ['oz', 'ounce'],

'pound': ['ibs', 'Ibs', 'lb', 'lbs', 'pound'],

'ton': ['ton', 't']

}

We are looking for these keywords in the OCR text. When found we extract the value and the unit and return the answer

For voltage - unit\_map = {

'kilovolt': ['kV', 'kilovolt', 'kv'],

'millivolt': ['mV', 'millivolt', 'mv'],

'volt': ['V', 'volt', 'v']

}

We are looking for these keywords in the OCR text. When found we extract the value and the unit and return the answer

For wattage - unit\_map = {

'kilowatt': ['kW', 'kilowatt', 'kw'],

'watt': ['W', 'watt', 'w']

}  
We are looking for these keywords in the OCR text. When found we extract the value and the unit and return the answer

For item\_volume - unit\_map = {

'centilitre': ['cl', 'centilitre'],

'cubic foot': ['cubic foot', 'ft³'],

'cubic inch': ['cubic inch', 'in³'],

'cup': ['cup'],

'decilitre': ['dl', 'decilitre'],

'fluid ounce': ['Fl','Fl 0z','fl oz','fl. oz','fi. oz', 'fluid ounce', 'oz'],

'gallon': ['gallon', 'gal'],

'imperial gallon': ['imperial gallon', 'imp gal'],

'litre': ['litre', 'liter', 'l'],

'microlitre': ['microlitre', 'µl'],

'millilitre': ['millilitre', 'ml'],

'pint': ['pint', 'pt'],

'quart': ['quart', 'qt']

}

We are looking for these keywords in the OCR text. When found we extract the value and the unit and return the answer

Due to our Limited processing power on our laptops, we use add ons to the downloading and processing function. We save intermediate (every 5000 images) results so that we don’t lose data in processing  
  
After attaining the partial results in multiple files (attached in the Zip), we use glob and python scripting to concatenate them into one csv  
  
Using more script we create the final submission.csv file.  
  
At the moment our program is unable to process Height Width and Depth, and we hope to add that in the next submission from our team.