Jingwen Zhong Strategies Write up

Read file: When I read file, I read each line one by one, and I separate the first part and second part of the line and go over all the words through **WordNetlemmatizer()**, then output them as "color" array and "object" array, and it will look like this.

```
1 color
```

```
['b', 'b', 'g', 'g', 'r', 'b', 'w']
```

```
1 objects
```

```
['ribbon', 'hair', 'hair', 'eye', 'lip', 'box', 'teeth']
```

Then I input "color" and "objects" in my function "predict".

In the function predict, I first call **Brown corpus** package, and go over each word and lemmatize it. Then I use **ngrams(N=2)** to pair 2 words together, then I call **FeqDist** to calculate each pair's frequency and it will return a dictionary "FeqDist" with the pair and its frequency.

Test: Now I know the frequency of each pair's in Brown corpu, I can use the inputs of the function ("color" and "objects" array). For each object in the "objects" array, I compare the second part of each pair in the "FeqDist" dictionary. If they are equal, then I check whether the object(a letter) in the "color" array (corresponding to the object in the "objects" array) exists in the first part of the pair, and also I use **matplot.color.css4_colors** and define a list called "color_set" that contains so many color's names, so I check if the first part of the pair in the "color_set". If they are satisfied, then I use these pairs and its frequency as candidates. If these two requirements are not met, I directly append this input (color(a letter) and object) into candidates set and also assign the frequency as 1. Then I change the candidates list set to dictionary and compare the frequency of each candidate, the one with the highest frequency is the result.

Bonus: By now, we can get the results, but the result might be the original input(color(a letter) and object). I loop each candidate, and if the first word(which is a letter) is not in the "color_set", then I check if this letter ever appears in some color names, and if it does, then I check if that color equals to the first letter of brown corpus pairs, and add up 1 to the frequency of that color every time when it appears in the corpus pair. Then I substitute the letter with the most frequent color. So far we can get a valid result of those who do not occur in the Brown corpus.