

```

import sys
import os

class preprocessing:
    def __init__(self, directory_name):
        self.directory = directory_name
        self.pos_directory = directory_name + "\\pos"
        self.neg_directory = directory_name + "\\neg"

    def run(self, output):
        self.pos_files = []
        for first, second, third in os.walk(self.pos_directory):
            for file in third:
                self.pos_files.append(file)
        self.neg_files = []
        for first, second, third in os.walk(self.neg_directory):
            for file in third:
                self.neg_files.append(file)
        self.pos = []
        self.neg = []

        for filename in self.pos_files:
            with open(self.pos_directory + "\\" + filename, "r", encoding="utf-8") as file:
                for line in file:
                    line = line.lower()
                    corrections = 0
                    for letter in range(len(line)):
                        if not line[letter+corrections].isalnum() and not
line[letter+corrections].isspace():
                            if(line[letter+corrections] != '\\' and line[letter+corrections] != '-'):
#remove
                                line = line[:letter+corrections] + " " + line[letter+corrections:]
                                corrections += 1
                            self.pos.append(line.split())

        for filename in self.neg_files:
            with open(self.neg_directory + "/" + filename, "r", encoding="utf-8") as file:
                for line in file:
                    line = line.lower()
                    corrections = 0
                    for letter in range(len(line)):
                        if not line[letter+corrections].isalnum() and not
line[letter+corrections].isspace():
                            line = line[:letter+corrections] + " " + line[letter+corrections:]
                            corrections += 1
                        self.neg.append(line.split())

        with open(output, "w", encoding="utf-8") as file:
            for line in self.pos:
                working_dict = {}
                output_string = ""
                for word in line:
                    try:
                        working_dict[word] += 1
                    except:
                        working_dict[word] = 1
                output_string += "+ "
                for key in working_dict:
                    output_string += key + " " + str(working_dict[key]) + " "
                output_string += "\n"
                file.write(output_string)

            for line in self.neg:
                working_dict = {}

```

```
output_string = ""
for word in line:
    try:
        working_dict[word] += 1
    except:
        working_dict[word] = 1
output_string += "- "
for key in working_dict:
    output_string += key + " " + str(working_dict[key]) + " "
output_string += "\n"
file.write(output_string)
```

```
if __name__ == "__main__":
    try:
        # first arg is directory name
        dir = sys.argv[1]
        # second arg is the output filename
        out = sys.argv[2]
        prep = preprocessing(dir)
        prep.run(out)
    except:
        print("missing pos/neg directory (arg 1) or missing output filename (arg 2)")
        print("try again please ...")
```