word_cloud.py

```python
# Fungkiat Phadejtaku
# 60070069
# github: https://github.com/CokeFung/introduction2NLP2020

#python3
#import nltk
#nltk.download("book")
from nltk.book import *
import os
from os import path
from wordcloud import WordCloud, STOPWORDS

def check_word(string):
    alp = "abcdefghijklmnopqrstuvwxyz".upper()
    if len(string) < 2:
        return False
    for i in alp:
        if i in string:
            return True
    return False

def list2string(word_list):
    text = " "
    for i in word_list:
        #print("%04d" % word_list[i], i)
        text += (i[0]+"#")*i[1]
    #print(text)
    return text

def string2wordcloud(text, num):
    wordcloud = WordCloud(max_words=30, max_font_size=70, scale=10, margin=5, stopwords=['#', '#']).generate(text)
    #crate directory before run code.
    wordcloud.to_file("./pic/wordcloud_text" + str(num) + ".png")
    # image = wordcloud.to_image()
    # image.show()

def writeFreq2File(text, num):
    text2write = "No.\tWord\t\tFrequency\n\n"
    for i in range(len(text)):
        text2write += ("%02d" % (i+1))+"\t"+text[i][0]+"\t\t"+str(text[i][1])+"\n"
    f = open(("./word_freq/word_freq_text"+str(num)+".txt"), "w+")
    f.write(text2write)
    f.close()

def doWordCloud(words_from_book, num):
    words_fd = FreqDist(words_from_book)
    words = {}
    for i in words_fd:
        #print(i, words_fd[i])
        tmp = i.upper()
        if check_word(tmp):
            if tmp in words:
                words[tmp] += words_fd[i]
            else:
                words[tmp] = words_fd[i]
    text = sorted(sorted(words.items()), key=lambda x:x[1]*-1)[:30]
    #print(text)
    text2 = list2string(text)
    string2wordcloud(text2, num)
    writeFreq2File(text, num)

def main():
    #generate text books's name
    for i in range(1, 10):
        doWordCloud(eval('text'+str(i)), i)

main()
```