from docx import Document  
x = 1  
able = []  
ablex = []  
def input(doc):  
 global x  
 gg = 0  
 able1 = []  
 yes = []  
 add = 0  
 ll = ""  
 for i in range(0, len(doc.paragraphs)):  
 b = (doc.paragraphs[i].text).split()  
 p = 0  
 o = 0  
 for j in range(0, len(b)):  
 k = b[j]  
 for n in range(0, len(k)):  
 if ("a" <= k[n] <= "z" or "A" <= k[n] <= "Z" or k[n] == "/"):  
 add += 1  
 if (add == len(k) and add > 1):  
 able1.append(k)  
 add = 0  
 yes.append(p)  
 else:  
 add = 0  
 p += 1  
 yes.append(-1)  
 for i in range(len(yes) - 1):  
 if (yes[i] + 1 == yes[i + 1]):  
 ll = ll + able1[o] + " "  
 o += 1  
 else:  
 ll = ll + able1[o]  
 o += 1  
 if(x==1):  
 able.append(ll)  
 else:  
 ablex.append(ll)  
 ll = ""  
 yes.clear()  
 able1.clear()  
 o = 0  
 if(x==2):  
 ableupdata1 = []  
 ableupdata2 = []  
 for i in able:  
 if i not in ableupdata1:  
 ableupdata1.append(i)  
 for i in ablex:  
 if i not in ableupdata2:  
 ableupdata2.append(i)  
 for i in range(0, len(ableupdata1)):  
 for j in range(0, len(ableupdata2)):  
 if (ableupdata1[i] ==ableupdata2[j]):  
 gg += 1  
 print("提取加过滤前的四级词汇有%d个单词或短语" % len(able))  
 print("提取加过滤前的高考词汇有%d个单词或短语" % len(ablex))  
 print("提取加过滤后的四级词汇有%d个单词或短语"%len(ableupdata1))  
 print("提取加过滤后的高考词汇有%d个单词或短语"%len(ableupdata2))  
 print("提取的四级单词或短语:")  
 print(ableupdata1)  
 print("提取的高考单词或短语:")  
 print(ableupdata2)  
 print("%d个元素相等" %gg)  
 print("检测到四级词汇和高考词汇的相似度为：%.2f%%"%((gg/len(ableupdata1))\*100))  
 print("执行结束")  
 else:  
 x += 1  
def main():  
 print("程序开始执行")  
 doc = Document('B://四级词汇.docx')  
 input(doc)  
 doc = Document('B://高中词汇.docx')  
 input(doc)  
 print("程序执行完毕")  
if \_\_name\_\_ == '\_\_main\_\_':  
 main()