

HW1 B10730012 張育瑋

2021/10/26

第一題

建立函示

In [1]:

```
with open("病歷500句子.txt", "r", encoding = "UTF-8") as f:
    f_content = f.read()
    vac = dict() #dictionary
    f.seek(0) #從頭讀取

    while len(f_content) > 0: #若還有文字就繼續讀取
        f_content = f.read(1)
        if f_content in vac.keys():
            vac[f_content] += 1
        else:
            vac[f_content] = 1

    #刪除雜項
    to_be_deleted = '\n:/》...; -、。· 5) ( '
    key_to_be_deleted = [key for key in vac if set(key).intersection(to_be_deleted) != set()]
    for key in key_to_be_deleted:
        vac.pop(key)
```

總字數

In [2]:

```
print("總共用了" + str(len(vac)) + " 個不同的字\n")
```

總共用了964 個不同的字

每字出現次數

In [3]:

```
print("每個字的出現次數")
for value in {value for value in vac.values()}:
    print(value)
    print({key for key in vac.keys() if vac[key] == value})
```

每個字的出現次數

```
1
{'', '瀟', '踏', '配', '糞', '寒', '褐', '據', '闌', '釋', '醒', '錯', '索',
'供', '勝', '塌', '鵲', '括', '氏', '臍', '膠', '站', '男', '疫', '痠', '島',
'準', '躁', '昨', '祕', '舊', '晨', '彬', '昏', '例', '當', '搏', '剖', '椅',
'貨', '垂', '桃', '沖', '肌', '濕', '桿', '櫃', '皓', '共', '汽', '正', '摘',
'睛', '應', '孕', '惡', '氧', '日', '居', '顏', '巧', '莖', '太', '效', '靈',
'酮', '撤', '屈', '伴', '磷', '鋼', '幹', '夾', '貧', '荀', '抑', '併', '葡',
'列', '肱', '付', '錄', '髓', '稍', '顆', '保', '鬆', '阿', '美', '振', '梯',
'充', '提', '並', '圖', '罩', '細', '蠟', '臨', '擔', '施', '卡', '末', '喝',
'拇', '統', '胞', '塑', '消', '畏', '取', '妄', '浸', '瓶', '盡', '顛', '扁',
'陌', '何', '陸', '盪', '凌', '極', '料', '偕', '強', '速', '迫', '沉', '顛',
'柯', '傳', '野', '象', '持', '釘', '討', '噴', '疏', '根', '盪', '限', '樹',
'妥', '快', '金', '調', '飆', '渣', '挫', '詳', '隱', '亂', '營', '講', '系',
'買', '古', '稱', '末', '巢', '聰', '借', '禁', '舌', '蒸', '資', '冷', '敷',
'伙', '耐', '癩', '箱', '寫', '灼', '顯', '匹', '旁', '癬', '黑', '訓', '格',
'冠', '癩', '木', '繃', '冒', '運', '隨', '砸', '緊', '醉', '談', '核', '胺',
'梢', '鐘', '洞', '澡', '卵', '免', '沙', '句', '萬', '陽', '散', '最', '寶',
'疣', '礪', '縫', '丁', '往', '堆', '褲', '乙', '學', '養', '千', '殖', '宗',
'停', '鈍', '臥', '增', '架', '震', '近', '確', '順', '割', '厲', '新', '寧',
'練', '線', '軟', '淳', '廁', '頂', '斯', '革', '嘗', '填', '尋', '遽', '份',
'撐', '產', '修', '套', '瞳', '餘', '暖', '友', '燥', '浴', '害', '叫', '輔',
'腋', '夠', '勃', '衡', '聲', '命', '淨', '數', '患', '裝', '零', '午', '潤',
'著', '論', '緒', '彈', '脾', '初', '齟', '總', '或', '算', '貉', '獨', '輪',
'任', '舌'}
```

```
2
{'漱', '唾', '硬', '噁', '驗', '減', '粥', '腮', '清', '角', '疾', '陪',
'頷', '置', '慈', '擴', '慾', '制', '淺', '事', '寶', '其', '像', '控', '克',
'嚙', '光', '嗎', '破', '剛', '璃', '燙', '單', '律', '希', '假', '鈣', '吻',
'嚴', '試', '趾', '書', '層', '維', '相', '必', '蘭', '容', '孔', '舒', '啡',
'圈', '險', '塊', '顧', '痘', '逆', '0', '憂', '嗆', '抬', '底', '鈉', '臼',
'濟', '樣', '面', '望', '斤', '收', '功', '防', '鎮', '吞', '問', '黃', '返',
'花', '優', '湯', '插', '遠', '絲', '升', '肘', '群', '陷', '星', '把', '巾',
'坐', '須', '醇', '漏', '紮', '偶', '易', '戴', '職', '嘴', '肚', '拒', '鎖',
'題', '纖', '立', '測', '補', '活', '方', '凝', '劇', '地', '述', '縱', '魚',
'於', '米', '雷', '唇', '什', '鐵', '海', '玻', '悶', '此', '偏', '代', '費',
'似', '林', '受', '麼', '乏'}
```

```
3
{'採', '針', '宮', '攝', '撕', '憩', '走', '股', '牛', '異', '臂', '實',
'深', '西', '北', '訂', '響', '移', '缺', '乘', '才', '癢', '非', '慢', '嗽',
'液', '淋', '荷', '齒', '只', '翻', '馬', '喉', '差', '丸', '助', '乾', '摔',
'創', '葉', '從', '明', '懷', '福', '鬱', '東', '推', '品', '交', '由', '鋸',
'副', '躺', '器', '片', '社', '支', '臍', '拿', '酸', '和', '風', '蠻', '彎',
'尾', '鉀', '七', '同', '餐', '脫', '薄', '百', '鮮', '毒', '棒', '鹽', '貼',
'窄', '暫', '曲', '踝', '員', '名', '囁', '汗'}
```

```
4
{'位', '羣', '辦', '=', '引', '騎', '待', '少', '靜', '狹', '痔', '計',
'灌', '衛', '脊', '放', '失', '影', '溝', '評', '乳', '智', '精', '晶', '退',
'倒', '聽', '係', '反', '模', '肥', '咳', '熱', '穩', '尖', '讓', '飲', '墊',
'裡', '足', '包', '屢', '台', '奶', '每', '肩', '半', '橫', '估', '劑', '息',
'周', '拍', '又', '瓣', '報', '適', '變', '滲', '越', '行', '穿', '菌', '掉',
'類'}
```

```
5
{'疝', '危', '輕', '室', '對', '臀', '條', '肉', '機', '等', '滴', '枕',
```

'油', '作', '整', '肝', '神', '拔', '組', '繼', '成', '音', '痰', '疑', '起',
'青', '己', '肪', '蜂', '張', '早', '溫', '平', '撞', '查', '腔', '期', '色',
'瘀', '而', '癒', '公', '織', '波', '法', '晚', '今', '路', '射', '至', '間',
'禍', '耳', '旦', '九'}
6
{ '冰', '約', '特', '膈', '眠', '裂', '麻', '照', '察', '致', '腰', '換',
'糊', '胰', '突', '送', '抗', '暈', '球', '帶', '觀', '阻', '想', '瘡', '困',
'需', '甲', '陰', '止', '巴', '洗', '潰', '固', '蓋', '我', '電', '指', '訴',
'吸', '難' }
7
{ '泌', '牙', '量', '障', '安', '脹', '床', '普', '很', '嘔', '染', '無',
'低', '離', '脂', '疹', '曾', '通', '素', '膜', '肋', '復', '呼', '膚', '刺',
'本', '常', '好', '泡', '束' }
8
{ '但', '得', '第', '椎', '微', '定', '被', '節', '別', '夜', '及', '窩',
'合', '重', '主', '長', '班', '住', '臉', '概', '注', '碎', '囊', '意', '告',
'預', '接', '教', '檢', '質', '也', '塞' }
9
{ '打', '抽', '睡', '糖', '物', '屬', '跌', '健', '覆', '視', '如', '吐',
'請', '膝', '隔', '鼻', '果', '症' }
10
{ '況', '工', '備', '排', '導', '們', '膿', '便', '蹤', '目', '癌', '腦',
'五', '話', '子', '膏', '能', '知', '腿', '全', '追', '度', '白', '求', '再' }
11
{ '布', '些', '拜', '鏡', '紗', '完', '覺', '輸', '關', '喘', '禮', '服',
'造', '敏', '腳', '疼', '始', '食', '道' }
12
{ '體', '車', '積', '肛', '力', '膀', '護', '理', '胱', '分', '續', '中',
'善' }
13
{ '八', '肢', '號', '多', '瘍', '身', '加', '流', '進', '回', '脈', '改',
'年', '先' }
14
{ '表', '直', '頸', '示', '較', '轉', '腎', '膽', '擦', '切' }
15
{ '皮', '看', '邊', '解', '感', '腺', '除', '治', '比', '四', '化', '家' }
16
{ '師', '瘤', '己', '動' }
145
{ '過' }
17
{ '處', '肺', '療', '超', '六' }
19
{ '臟', '兩', '議', '胃', '狀', '背', '都', '建' }
20
{ '入', '頭' }
21
{ '幫', '月', '內' }
22
{ '小', '術' }
18
{ '外', '燒', '紅', '天', '房', '折' }
23
{ '然', '經', '點', '那', '候' }
24
{ '吃', '斷', '眼' }
26
{ '現', '壓', '雙' }
27
{ '自', '氣', '使', '腸', '手' }
28

```
{ '生', '部', '管', '科' }  
157  
{ '他' }  
29  
{ '石', '二' }  
31  
{ '門', '時', '情' }  
160  
{ '的' }  
33  
{ '炎', '上', '醫', '給' }  
32  
{ '大', '腹', '性', '三' }  
35  
{ '在', '可' }  
36  
{ '院', '前' }  
34  
{ '了', '個', '下', '之', '所' }  
38  
{ '口', '出' }  
161  
{ '病' }  
37  
{ '側', '骨' }  
41  
{ '發', '藥', '腫' }  
42  
{ '十' }  
43  
{ '傷', '不', '用' }  
44  
{ '尿', '史' }  
169  
{ '是' }  
40  
{ '跟' }  
48  
{ '要' }  
49  
{ '來' }  
51  
{ '還' }  
52  
{ '左' }  
53  
{ '沒', '次', '說', '到' }  
54  
{ '這' }  
58  
{ '痛' }  
59  
{ '以' }  
63  
{ '做', '去' }  
67  
{ '會' }  
69  
{ '一' }  
72  
{ '診' }
```

```
73
{'血'}
77
{'刀'}
79
{'後'}
337
{'有'}
89
{'為'}
90
{'右'}
93
{'人', '因'}
107
{'開'}
30
{'結', '高', '水', '心'}
120
{'就'}
25
{'胸', '急', '形'}
```

前十常出現的字及次數

In [4]:

```
print("前十常出現的字和其出現次數")
ke=list(vac.keys())
rank_sort_key=sorted(ke,key= lambda ke:vac[ke],reverse=True)

for i in rank_sort_key[:10]:
    print('{},{}'.format(i,vac[i]))
```

前十常出現的字和其出現次數

```
有,337
是,169
病,161
的,160
他,157
過,145
就,120
開,107
人,93
因,93
```

只出現過一次的字

In [5]:

```
print("只出現一次的字\n")
print({key for key in vac.keys() if vac[key] == 1})
```

只出現一次的字

```
{',', '瀟', '踏', '配', '糞', '寒', '褐', '據', '闌', '釋', '醒', '錯', '索',
'供', '勝', '塌', '鵲', '括', '氏', '臍', '膠', '站', '男', '疫', '痠', '島',
'準', '躁', '昨', '祕', '舊', '晨', '彬', '昏', '例', '當', '搏', '剖', '椅',
'貨', '垂', '桃', '沖', '肌', '濕', '桿', '櫃', '皓', '共', '汽', '正', '摘',
'睛', '應', '孕', '惡', '氧', '日', '居', '顏', '巧', '莖', '太', '效', '靈',
'酮', '撤', '屈', '伴', '磷', '鋼', '幹', '夾', '貧', '萄', '抑', '併', '葡',
'列', '肱', '付', '錄', '髓', '稍', '顆', '保', '鬆', '阿', '美', '振', '梯',
'充', '提', '並', '圖', '罩', '細', '蠟', '臨', '擔', '施', '卡', '末', '喝',
'拇', '統', '胞', '塑', '消', '畏', '取', '妄', '浸', '瓶', '盡', '顛', '扁',
'陌', '何', '陸', '盪', '凌', '極', '料', '偕', '強', '速', '迫', '沉', '顛',
'柯', '傳', '野', '象', '持', '釘', '討', '噴', '疏', '根', '盪', '限', '樹',
'妥', '快', '金', '調', '飄', '渣', '挫', '詳', '隱', '亂', '營', '講', '系',
'買', '古', '稱', '未', '巢', '聰', '借', '禁', '舌', '蒸', '資', '冷', '敷',
'伙', '耐', '癩', '箱', '寫', '灼', '顯', '匹', '旁', '癬', '黑', '訓', '格',
'冠', '癩', '木', '繃', '冒', '運', '隨', '砸', '緊', '醉', '談', '核', '胺',
'梢', '鐘', '洞', '澡', '卵', '免', '沙', '句', '萬', '陽', '散', '最', '寶',
'疣', '磺', '縫', '丁', '往', '堆', '褲', '乙', '學', '養', '千', '殖', '宗',
'停', '鈍', '臥', '增', '架', '震', '近', '確', '順', '割', '厲', '新', '寧',
'練', '線', '軟', '淳', '廁', '頂', '斯', '革', '嘗', '填', '尋', '遽', '份',
'撐', '產', '修', '套', '瞳', '餘', '暖', '友', '燥', '溶', '害', '叫', '輔',
'腋', '夠', '勃', '衡', '聲', '命', '淨', '數', '患', '裝', '零', '午', '潤',
'著', '論', '緒', '彈', '脾', '初', '齟', '總', '或', '算', '賂', '獨', '輪',
'任', '否'}
```

第二題

建立函式

In [16]:

```

def N_gram(file_name, N):
    with open(file_name, "r", encoding = "UTF-8") as f:
        f_content = f.read()
        n_dic = dict() #n-gram dictionary
        n_minus1_dic = dict() #(n-1)-gram dictionary
        pr = dict()

        f.seek(0) #從頭讀取

        #消除雜項
        f_content = f_content.replace(' ', '')
        f_content = f_content.replace('。', '')
        f_content = f_content.replace('。', '')
        f_content = f_content.replace('=', '')
        f_content = f_content.replace('\n', '')
        f_content = f_content.replace('、', '')
        f_content = f_content.replace('5', '')
        f_content = f_content.replace('0', '')

        for c in range(len(f_content)-N+1): #讀取N-gram出現次數
            if f_content[c:c+N] in n_dic.keys():
                n_dic[f_content[c:c+N]] += 1
            else:
                n_dic[f_content[c:c+N]] = 1

        for c in range(len(f_content)-(N-1)+1): #讀取(N-1)-gram出現次數
            if f_content[c:c+N-1] in n_minus1_dic.keys():
                n_minus1_dic[f_content[c:c+N-1]] += 1
            else:
                n_minus1_dic[f_content[c:c+N-1]] = 1

        #計算機率
        for c in range(len(f_content)-N): #讀取(N-1)-gram出現次數
            x = n_dic[f_content[c:(c+N)]] / n_minus1_dic[f_content[c:(c+N-1)]]
            pr[f_content[c:(c+N)]] = x

        print("=====")
        print("機率前十大的詞")
        ke=list(pr.keys())
        rank_sort_key=sorted(ke,key= lambda ke:pr[ke],reverse=True)

        for i in rank_sort_key[:10]:
            print('{},{}'.format(i,pr[i]))

        #
        # print("=====")
        # print("N-gram詞的出現機率")
        # for value in {value for value in pr.values()}:
        #     print(value)
        #     print({key for key in pr.keys() if pr[key] == value})

        pr.clear

        f.close()

```

列出2-gram機率

In [17]:

```
print('N-gram = 2\n')
N_gram("病歷500句子.txt", 2)
```

N-gram = 2

```
=====
機率前十大的詞
系統,1.0
統摘,1.0
摘錄,1.0
錄句,1.0
句子,1.0
什麼,1.0
建議,1.0
息肉,1.0
東西,1.0
米湯,1.0
```

列出3-gram機率

In [18]:

```
print('N-gram = 3\n')
N_gram("病歷500句子.txt", 3)
```

N-gram = 3

```
=====
機率前十大的詞
衛福部,1.0
福部台,1.0
部台北,1.0
北醫院,1.0
院護理,1.0
理交班,1.0
班系統,1.0
系統摘,1.0
統摘錄,1.0
摘錄句,1.0
```

列出4-gram機率

In [19]:

```
print('N-gram = 4\n')  
N_gram("病歷500句子.txt", 4)
```

N-gram = 4

```
=====
```

機率前十大的詞

衛福部台	1.0
福部台北	1.0
部台北醫	1.0
台北醫院	1.0
醫院護理	1.0
院護理交	1.0
護理交班	1.0
理交班系	1.0
交班系統	1.0
班系統摘	1.0

第三題

該字為 x_1 時，可以 100% 斷定其下一個字 x_2 出現的字

In [10]:

```

with open("病歷500句子.txt", "r", encoding = "UTF-8") as f:
    f_content = f.read()
    n_dic = dict() #n-gram dictionary
    n_minus1_dic = dict() #(n-1)-gram dictionary
    pr = dict()
    f.seek(0) #從頭讀取
    N = 2

    #消除雜項
    f_content = f_content.replace(' ', '')
    f_content = f_content.replace('。', '')
    f_content = f_content.replace('，', '')
    f_content = f_content.replace('=', '')
    f_content = f_content.replace('\n', '')
    f_content = f_content.replace('、', '')
    f_content = f_content.replace('5', '')
    f_content = f_content.replace('0', '')

    for c in range(len(f_content)-N+1): #讀取N-gram出現次數
        if f_content[c:c+N] in n_dic.keys():
            n_dic[f_content[c:c+N]] += 1
        else:
            n_dic[f_content[c:c+N]] = 1

    for c in range(len(f_content)-(N-1)+1): #讀取(N-1)-gram出現次數
        if f_content[c:c+N-1] in n_minus1_dic.keys():
            n_minus1_dic[f_content[c:c+N-1]] += 1
        else:
            n_minus1_dic[f_content[c:c+N-1]] = 1

    #計算機率
    for c in range(len(f_content)-N): #讀取(N-1)-gram出現次數
        x = n_dic[f_content[c:(c+N)]] / n_minus1_dic[f_content[c:(c+N-1)]]
        pr[f_content[c:(c+N)]] = x

    print("該字為 x1 時，可以 100% 斷定其下一個字 x2 出現的字。\\n")
    print({key[0] for key in pr.keys() if pr[key] == 1.0})

    pr.clear
    n_dic.clear
    n_minus1_dic.clear

    f.close()

```

該字為 x1 時，可以 100% 斷定其下一個字 x2 出現的字。

```

{'噁', '醒', '塌', '膠', '痠', '島', '事', '當', '寶', '貨', '控', '沖',
'濕', '桿', '皓', '共', '正', '睛', '報', '氣', '日', '希', '居', '巧', '莖',
'效', '撤', '鋼', '幹', '貧', '併', '葡', '舒', '稍', '穎', '提', '圖', '罩',
'細', '蠟', '抬', '碎', '取', '浸', '瓶', '盡', '扁', '陸', '凌', '偕', '強',
'速', '柯', '釘', '盥', '限', '樹', '調', '少', '品', '渣', '隱', '脊', '古',
'組', '伙', '箱', '癩', '寫', '匹', '砸', '洞', '卵', '沙', '縫', '肚', '褲',
'疝', '養', '停', '鈍', '增', '架', '割', '評', '薄', '厲', '斯', '嘗', '瞳',
'蜂', '海', '友', '害', '勃', '聲', '零', '息', '午', '論', '彈', '總', '漱',
'踏', '硬', '寒', '錯', '疾', '供', '慈', '擴', '鵲', '氏', '疫', '肥', '彬',
'例', '椅', '垂', '桃', '嗎', '櫃', '摘', '應', '孕', '顏', '太', '吻', '伴',

```

'磷', '慢', '夾', '容', '付', '髓', '阿', '淋', '美', '憂', '臨', '施', '卡',
'拇', '喝', '胞', '妄', '喉', '困', '陌', '功', '料', '顧', '野', '討', '懷',
'疏', '飄', '亂', '營', '講', '系', '買', '未', '繼', '晶', '蒸', '資', '灼',
'談', '胺', '偶', '梢', '癒', '陽', '散', '纖', '千', '臥', '痔', '確', '順',
'練', '目', '革', '遽', '產', '暖', '夠', '淨', '甲', '潤', '齟', '或', '算',
'任', '濾', '褐', '闌', '釋', '腮', '角', '頷', '勝', '憩', '髒', '牛', '男',
'準', '祕', '舊', '禮', '昏', '剖', '止', '嚙', '汽', '織', '惡', '律', '靈',
'屈', '嚴', '特', '苟', '抑', '保', '充', '並', '只', '擔', '統', '消', '濟',
'何', '極', '吞', '問', '迫', '傳', '象', '持', '束', '需', '根', '快', '詳',
'巢', '禁', '舌', '音', '星', '模', '冷', '數', '耐', '癬', '黑', '格', '木',
'使', '隨', '緊', '醉', '核', '鐘', '澡', '我', '免', '復', '礪', '往', '丁',
'堆', '鎖', '乙', '震', '近', '寧', '縱', '軟', '淳', '米', '雷', '尋', '份',
'什', '撐', '修', '套', '餘', '燥', '叫', '輔', '命', '患', '裝', '緒', '潰',
'初', '獨', '隴', '乏', '唾', '辦', '配', '糞', '膀', '據', '索', '撕', '括',
'股', '站', '昨', '躁', '晨', '觀', '搏', '其', '肌', '璃', '燙', '列', '肱',
'錄', '鬆', '建', '險', '振', '荷', '梯', '嘔', '末', '飲', '塑', '畏', '翻',
'顚', '盪', '鎮', '追', '沉', '噴', '妥', '金', '挫', '升', '稱', '聰', '借',
'顯', '旁', '阻', '訓', '冠', '癰', '酮', '繃', '冒', '運', '職', '勾', '萬',
'竇', '最', '疣', '嘴', '學', '殖', '宗', '機', '新', '線', '廁', '頂', '填',
'尖', '溶', '玻', '腋', '衡', '暫', '數', '著', '脾', '髁', '似', '輪', '否'}

第四題

建立函式

In [11]:

```
!pip install pypinyin
```

Requirement already satisfied: pypinyin in c:\users\user\anaconda3\lib\site-packages (0.43.0)

In [12]:

```

from pypinyin import pinyin
import os
os.system("cls")

with open("病歷500句子.txt", "r", encoding = "UTF-8") as f:
    f_content = f.read()
    pinyin_count = dict() #n-gram dictionary

    f.seek(0) #從頭讀取

    #消除雜項
    f_content = f_content.replace(' ', '')
    f_content = f_content.replace('。', '')
    f_content = f_content.replace('。', '')
    f_content = f_content.replace('=', '')
    f_content = f_content.replace('\n', '')
    f_content = f_content.replace('、', '')
    f_content = f_content.replace('5', '')

    for c in range(len(f_content)): #讀取N-gram出現次數
        if str(pinyin(f_content[c])) in pinyin_count.keys():
            pinyin_count[str(pinyin(f_content[c]))] += 1
        else:
            pinyin_count[str(pinyin(f_content[c]))] = 1

    f.close()

```

所有字一共有幾種讀音

In [13]:

```

print("=====")
print("所有字一共有幾種讀音")
ke=list(pinyin_count.keys())
print(len(ke))

```

```

=====
所有字一共有幾種讀音
610

```

出現最多次的讀音和他的出現次數

In [14]:

```

print("=====")
print("出現最多次的讀音和他的出現次數")

rank_sort_key=sorted(ke, key= lambda ke:pinyin_count[ke], reverse=True)

for i in rank_sort_key[:1]:
    print('{}{}'.format(i, pinyin_count[i]))

```

```

=====
出現最多次的讀音和他的出現次數
[['yǒu']], 338

```

第五題

In [15]:

```

from pypinyin import pinyin

with open("病歷500句子.txt", "r", encoding = "UTF-8") as f:
    f_content = f.read()
    word_to_pinyin = {} #n-gram dictionary

    f.seek(0) #從頭讀取

    #消除雜項
    f_content = f_content.replace(' ', '')
    f_content = f_content.replace('。', '')
    f_content = f_content.replace('．', '')
    f_content = f_content.replace('＝', '')
    f_content = f_content.replace('\n', '')
    f_content = f_content.replace('、', '')
    f_content = f_content.replace('5', '')
    f_content = f_content.replace('0', '')

    #建立[文字:拼音]的dict
    for c in range(len(f_content)-1): #讀取N-gram出現次數
        if f_content[c:c+2] not in word_to_pinyin.keys():
            word_to_pinyin[f_content[c:c+2]] = str(pinyin(f_content[c:c+2]))

    print("=====")
    print("同音異字者")

    for value in {value for value in word_to_pinyin.values()}:

        same_pinyin = [] #建立同音的list
        for key in word_to_pinyin.keys(): #若同音就加入list
            if str(word_to_pinyin[key]) == value:
                same_pinyin.append(key)

        if len(same_pinyin)>1: #印出list內容超過一個的同音詞
            print(value)
            print(same_pinyin)

    f.close()

```

```

=====
同音異字者
[['jiān'], ['huì']]
['間會', '肩會']
[['zhì'], ['yīn']]
['致陰', '治因']
[['hòu'], ['tā']]
['候他', '後他']
[['zhěn'], ['yǒu']]
['疹有', '診有']
[['shì'], ['bìng']]
['是病', '示病', '視病']
[['shì'], ['zì']]
['是自', '示自']
[['shì'], ['dào']]
['是到', '適到']
[['tòng'], ['yòu']]
['痛又', '痛右']
[['yǒu'], ['jìn']]
['有浸', '有進']

```

[['hòu'], ['shì']]
['後是', '候是']
[['shì'], ['zhǐ']]
['是只', '是止']
[['dào'], ['bù']]
['道不', '到不']
[['mài'], ['qū']]
['脈屈', '脈曲']
[['shù'], ['zhè']]
['術這', '束這']
[['zhì'], ['bìng']]
['質並', '質病']
[['shù'], ['fù']]
['束腹', '述腹']
[['yǒu'], ['jiào']]
['有叫', '有教']
[['shì'], ['tā']]
['釋他', '是他']
[['niào'], ['bù']]
['尿布', '尿不']
[['yào'], ['jiù']]
['要就', '藥就']
[['chá'], ['zhè']]
['察這', '查這']
[['hòu'], ['yǒu']]
['候有', '後有']
[['zuò'], ['chù']]
['做處', '作處']
[['dāo'], ['yī']]
['刀醫', '刀一']
[['wán'], ['zhī']]
['完支', '完之']
[['chāo'], ['yīn']]
['超因', '超音']
[['xiàn'], ['qíng']]
['陷情', '限情']
[['yào'], ['kě']]
['要可', '藥可']
[['xìng'], ['gēn']]
['性跟', '性根']
[['shí'], ['yǒu']]
['十有', '石有', '食有', '時有']
[['dào'], ['jiā']]
['到加', '到家']
[['wèi'], ['yào']]
['胃藥', '為藥', '為要']
[['shì'], ['qián']]
['是前', '是乾']
[['bàn'], ['zuò']]
['半做', '半坐']
[['yǒu'], ['bàn']]
['有辦', '有伴']
[['shī'], ['yǒu']]
['師有', '施有']
[['cháng'], ['zhōng']]
['常腫', '腸腫']
[['gōng'], ['xuè']]
['供血', '工血']
[['shì'], ['yī']]
['是一', '是醫', '室一']
[['nián'], ['yī']]

['年醫', '年一']
[['wèi'], ['fù']]
['為腹', '為副']
[['tā'], ['xiàn']]
['塌陷', '他現']
[['yào'], ['de']]
['藥的', '要的']
[['yǐ'], ['sān']]
['以三', '已三']
[['zhǒng'], ['liú']]
['腫瘤', '腫流']
[['shì'], ['shǎo']]
['試少', '是少']
[['tā'], ['zhù']]
['他注', '他助']
[['zhào'], ['gù']]
['照顧', '罩固']
[['gǔ'], ['kē']]
['骨科', '古柯']
[['jiào'], ['tā']]
['叫他', '教他']
[['shì'], ['shì']]
['似是', '示是']
[['hòu'], ['zuǒ']]
['後左', '候左']
[['bèi'], ['cā']]
['被擦', '背擦']
[['gé'], ['lán']]
['革蘭', '格蘭']
[['bù'], ['gēn']]
['布跟', '部跟']
[['shí'], ['gēn']]
['石跟', '食跟']
[['shí'], ['yòu']]
['時又', '石右']
[['shì'], ['shuō']]
['是說', '示說']
[['zhǐ'], ['yǒu']]
['只有', '趾有', '指有']
[['dào'], ['hòu']]
['到後', '道後']
[['shí'], ['hòu']]
['時候', '時後']
[['yòng'], ['yī']]
['用一', '用醫']
[['qì'], ['guǎn']]
['氣管', '汽管']
[['shì'], ['jīn']]
['是今', '示今', '是金']
[['shì'], ['yòu']]
['是右', '示右']
[['shì'], ['yǒu']]
['是有', '試有']
[['yào'], ['zài']]
['要在', '藥在', '要再']
[['de'], ['bù']]
['的部', '的不']
[['de'], ['yì']]
['的罩', '的異']
[['jīng'], ['yī']]
['經一', '經醫']

[['shì'], ['xiān']]
['是先', '示先']
[['zhěn'], ['jīng']]
['診經', '診精']
[['yǒu'], ['zhī']]
['有脂', '有支']
[['yào'], ['yǒu']]
['藥有', '要有']
[['qū'], ['zhāng']]
['屈張', '曲張']
[['tā'], ['jù']]
['他拒', '他鋸']
[['shì'], ['yīn']]
['是因', '氏陰']
[['jiā'], ['jù']]
['加劇', '加遽']
[['rén'], ['gāng']]
['人肛', '人剛']
[['shí'], ['wèi']]
['食衛', '石未']
[['de'], ['huà']]
['的話', '的化']
[['shì'], ['yào']]
['是要', '示要']
[['de'], ['yī']]
['的一', '的醫']
[['yǒu'], ['yào']]
['有要', '有藥']
[['yǒu'], ['shèn']]
['有滲', '有腎']
[['yòng'], ['shù']]
['用束', '用術']
[['hòu'], ['bèi']]
['後背', '候被']
[['shì'], ['jiàn']]
['是建', '適建']
[['hòu'], ['huì']]
['後會', '候會']
[['tā'], ['dài']]
['他戴', '他待']
[['hòu'], ['yào']]
['候要', '後藥']
[['xī'], ['jiù']]
['西就', '膝就']
[['jù'], ['dào']]
['劇到', '鋸到']
[['hòu'], ['bìng']]
['候病', '後病']
[['shí'], ['dào']]
['十到', '食道']
[['yǒu'], ['xiān']]
['有先', '有鮮']
[['yì'], ['bìng']]
['意病', '議病']
[['hòu'], ['jiù']]
['後就', '候就']
[['hòu'], ['xuè']]
['候血', '後血']
[['dào'], ['shí']]
['到十', '到時']
[['dài'], ['shǐ']]

```
['待使', '帶使']  
[['yǒu'], ['qì']]  
['有氣', '有憩']  
[['wèi'], ['bèi']]  
['為背', '為被']  
[['dào'], ['jiù']]  
['到就', '道就']  
[['wèi'], ['zhì']]  
['為痔', '位置']  
[['shì'], ['qù']]  
['是去', '適去']  
[['liáo'], ['yào']]  
['療藥', '療要']  
[['yǒu'], ['xīn']]  
['有心', '有新']  
[['de'], ['wèi']]  
['的胃', '的位']  
[['shǐ'], ['yǒu']]  
['始有', '史有']  
[['guò'], ['zhī']]  
['過之', '過脂']  
[['qián'], ['yào']]  
['前要', '前藥']  
[['jiān'], ['yǒu']]  
['肩有', '間有']  
[['yǒu'], ['zài']]  
['有再', '有在']
```

In []: