

Allion Test Labs, Inc.

Translation Module User Guide

Version 1.0.0

2019/06/17

Revision History

Version	Revise Date	Author	Revise Note	Revise By
V1.0	2019-06-17	Rex	Initial	Rex Wu

此 Module comepile 為 exe，執行 translation_module.exe -h 顯示說明：

```
usage: translation_module.exe [-h] -p P -r R [-lang LANG]

optional arguments:
  -h, --help      show this help message and exit
  -p P            json file path of translation list
  -r R            path of result json
  -lang LANG      List of available languages: afr, amh, ara, asm, aze, aze_cyrl, bel, ben,
                  bod, bos, bul, cat, ceb, ces, chi_sim, chi_tra, chr, cym, dan, dan_frak, deu,
                  deu_frak, dzo, ell, eng, enm, epo, equ, est, eus, fas, fin, fra, frk, frm, glg,
                  grc, guj, hat, heb, hin, hrv, hun, iku, ind, isl, ita, ita_old, jav, jpn, kan,
                  kat, kat_old, kaz, khm, kir, kor, kur, lao, lat, lav, lit, mal, mar, mlt, msa,
                  mya, nep, nld, nor, ori, osd, pan, pol, por, pus, ron, rus, san, sin, slk,
                  slk_frak, slv, spa, spa_old, sqi, srp, srp_latn, swa, swe, syr, tam, tel,
                  tgk, tgl, tha, tir, tur, uig, ukr, urd, uzb, uzb_cyrl, vie, yid
```

1 參數 p：要翻譯的圖片列表路徑,存以 json 格式.

- 最上層的鍵值(此以數字為例),為不可重複的 key,產出的 result 會以此鍵值對應。
- answer 為需要驗證的字串, 經 OCR 後會比對此字串與結果之差異。如不需驗證請留空字串,則直接產出 OCR 結果
- pic 為圖片的完整路徑。

```
{
  "0": {
    "answer": "אשור",
    "pic": "Z:/TV/Stage1/Translation_cs1-2_SQL-U3-SG-CVT(AEP)-43-91-MB-50"
  },
  "1": {
    "answer": "בטוח",
    "pic": "Z:/TV/Stage1/Translation_cs1-2_SQL-U3-SG-CVT(AEP)-43-91-MB-50"
  },
  "2": {
    "answer": "Күшін жою",
    "pic": "Z:/TV/Stage1/Translation_cs1-2_SQL-U3-SG-CVT(AEP)-43-91-MB-50"
  },
  "3": {
    "answer": "Жок",
    "pic": "Z:/TV/Stage1/Translation_cs1-2_SQL-U3-SG-CVT(AEP)-43-91-MB-50"
  },
  "4": {
    "answer": "Atcelt",
    "pic": "Z:/TV/Stage1/Translation_cs1-2_SQL-U3-SG-CVT(AEP)-43-91-MB-50"
  },
  "5": {
    "answer": "",
    "pic": "Z:/TV/Stage1/Translation_cs1-2_SQL-U3-SG-CVT(AEP)-43-91-MB-50"
  }
}
```

2 參數 r 為產出 result 的資料夾

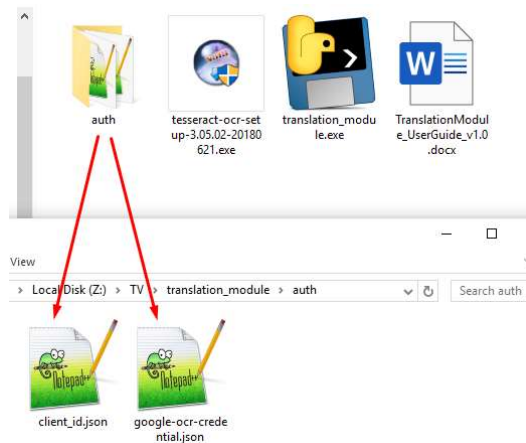
- 檔名以 timestamp 加註: result_20190617152715.json

3 參數 lang 為指定 OCR 的語言

- 如下此參數則以 tesseract 為 ocr engine
- 需安裝 tesseract-ocr-setup-3.05.02-20180621.exe
- 支援的語言列表:
 afr,amh,ara,asm,aze,aze_cyrl,bel,ben,bod,bos,bul,cat,ceb,ces,chi_sim,chi_tra,chr,cym,dan,dan_frak,deu,deu_frak,dzo,ell,eng,enm,epo,equ,est,eus,fas,fin,fra,frk,frm,gle,glg,grc,guj,hat,heb,hin,hrv,hun,iku,ind,isl,ita,ita_old,jav,jpn,kan,kat,kat_old,kaz,khm,kir,kor,kur,lao,lat,lav,lit,mal,mar,mkd,mlt,msa,mya,nep,nld,nor,ori,osd,pan,pol,popr,pus,ron,rus,san,sin,slk,slk_frak,slv,spa,spa_old,sqi,srp,srp_latn,swa,swe,syr,tam,tel,tgk,tgl,tha,tir,tur,uig,ukr,urd,uzb,uzb_cyrl,vie,yid

4 以 Google Drive 為 OCR engine

- 如未指定 lang 參數則採 Google Drive 為 OCR engine
- 在 module 同目錄下需有 auth folder, 內含登入 Google drive 所需的驗證檔



- 使用環境須可連外網

5 Module 使用示例

- translation_module.exe -p "c:/translation_list.json" -r "z:/"

```

Z:\TV
λ translation_module.exe -p "c:/translation_list.json" -r "z:/"
C:\Users\REXWU1~1\AppData\Local\Temp\MEI157882\fuzzywuzzy\ fuzz.py:11: UserWarning: Using slow pure-python SequenceMatcher. Install python-Levenshtein to remove this warning
100% | 14/14 [02:24<00:00, 9.69s/it]
Result file : z:/result_20190617153358.json
  
```

產出的result路徑

6 Results

```
{
  "0": {
    "answer": "אישור",
    "pic": "Z:/TV/Stage1/Translation_cs1-2_SQL-U3-SG-CVT(AEP)-43-91-MB-50_v8.847_Kenny_2019-05-29/EU/hbr/Sound_Reset_dialogScreen_Cancel",
    "result": "pass",
    "course": "",
    "ocr": "אישור"
  },
  "1": {
    "answer": "כניסה",
    "pic": "Z:/TV/Stage2/Translation_cs1-2_SQL-U3-SG-CVT(AEP)-43-91-MB-50_v8.847_Kenny_2019-05-29/EU/hbr/Sound_Reset_dialogScreen_Cancel",
    "result": "fail",
    "course": "file not found",
    "ocr": ""
  },
  "2": {
    "answer": "Күшім жою",
    "pic": "Z:/TV/Stage1/Translation_cs1-2_SQL-U3-SG-CVT(AEP)-43-91-MB-50_v8.847_Kenny_2019-05-29/EU/kaz/Sound_Reset_dialogScreen_Cancel",
    "result": "fail",
    "course": "case not match",
    "ocr": "КҮШІМ ЖОЮ"
  },
  "3": {
    "answer": "Anuluj",
    "pic": "Z:/TV/Stage1/Translation_cs1-2_SQL-U3-SG-CVT(AEP)-43-91-MB-50_v8.847_Kenny_2019-05-29/EU/pol/Sound_Reset_dialogScreen_Cancel",
    "result": "fail",
    "course": "match 83%",
    "ocr": "Anuluj"
  },
  "4": {
    "answer": "",
    "pic": "Z:/TV/Stage1/Translation_cs1-2_SQL-U3-SG-CVT(AEP)-43-91-MB-50_v8.847_Kenny_2019-05-29/EU/lav/Sound_Reset_dialogScreen_Cancel",
    "result": "",
    "course": "",
    "ocr": "_____\\n\\nMV16-508 CHH JP\\nSD\\nAtjaunot\\nU PHOSA UMC1829\\nMOM\\n-0\\nVai tiešām vēlaties atjaunot rūpnieciskos iestādījumus?"
  },
  "5": {
    "answer": "Taisnība",
    "pic": "Z:/TV/Stage1/Translation_cs1-2_SQL-U3-SG-CVT(AEP)-43-91-MB-50_v8.847_Kenny_2019-05-29/EU/lav/Sound_Reset_dialogScreen_Cancel",
    "result": "pass",
    "course": "",
    "ocr": "Taisnība"
  }
}
```

answer與ocr結果一致且沒有重複

圖片路徑不正確

answer與ocr結果一致但大小寫不符

answer與ocr結果符合比例

answer為空則直接輸出完整ocr結果