Deidentification Class Python File Similarity Check

Abstract

This lab is for checking similarity between two python files which were created by BMI500 students in the Deidentification Assignment by using Levenshtein, Tensorflow, and STS Benchmark.

Methods

- Before doing all methods, all python files are preprocessed by neglecting stat.py because this file is commonly used by all students.

- Levenshtein

    - Jaro
    - Distance
    - Ration
- Tensorflow
- STS Benchmark

Results

Levenshtein

```
 1   Parisa deid.py : Shaojun deid.py
 2
 3   The similarity results tested by Levenshtein are following:
 4   jaro: 1.0
 5   distance: 0
 6   ratio: 1.0
 7
 8   Parisa deid.py : Shaojun deid-Shaojun Yu.py
 9
10   The similarity results tested by Levenshtein are following:
```

```
jaro: 0.6622267258095486
distance: 6217
ratio: 0.49563492063492065

Parisa deid.py : FeverDreams deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.6684480457968126
distance: 5400
ratio: 0.6632829908393166

Parisa deid.py : FeverDreams deid_phone.py

The similarity results tested by Levenshtein are following:
jaro: 0.822566622827735
distance: 261
ratio: 0.9741347227391143

Parisa deid.py : Itoshikihiro deid-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 1.0
distance: 0
ratio: 1.0

Parisa deid.py : Itoshikihiro deid.py

The similarity results tested by Levenshtein are following:
jaro: 1.0
distance: 0
ratio: 1.0

Parisa deid.py : Itoshikihiro deid-date-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 0.6557711323609124
distance: 2317
ratio: 0.6839400428265524

Parisa deid.py : Itoshikihiro deid-name-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 0.7251300664288114
distance: 2450
ratio: 0.6850410722006053

Parisa deid.py : Nasim de-id-location.py

The similarity results tested by Levenshtein are following:
jaro: 0.7288723120536037
distance: 3611
ratio: 0.6504225995583645

```

```
Parisa deid.py : Nasim deid.py

The similarity results tested by Levenshtein are following:
jaro: 1.0
distance: 0
ratio: 1.0

Parisa deid.py : Nasim de-id-date.py

The similarity results tested by Levenshtein are following:
jaro: 0.7532462007044755
distance: 2080
ratio: 0.8198847262247838

Parisa deid.py : Nasim de-id-age.py

The similarity results tested by Levenshtein are following:
jaro: 0.779738456966431
distance: 2053
ratio: 0.7371695178849145

Parisa deid.py : Andre deid-Andre_Holder.py

The similarity results tested by Levenshtein are following:
jaro: 0.7873682872182086
distance: 1287
ratio: 0.8806998939554613

Parisa deid.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.8231258743340041
distance: 211
ratio: 0.9756950837783097

Parisa deid.py : Mahmoud deid_date_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7991964384755024
distance: 702
ratio: 0.9272179918540818

Parisa deid.py : Mahmoud deid_age_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.8142641234970611
distance: 427
ratio: 0.950575859254557

Parisa deid.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.5947612031278235
```

```
117   distance: 14970
118   ratio: 0.41918212151780865
119
120   Parisa deid.py : ChenLin deid.py
121
122   The similarity results tested by Levenshtein are following:
123   jaro: 1.0
124   distance: 0
125   ratio: 1.0
126
127   Parisa deid_Date_Parisa.py : Shaojun deid.py
128
129   The similarity results tested by Levenshtein are following:
130   jaro: 0.7322073315167268
131   distance: 2398
132   ratio: 0.8149387693009812
133
134   Parisa deid_Date_Parisa.py : Shaojun deid-Shaojun Yu.py
135
136   The similarity results tested by Levenshtein are following:
137   jaro: 0.7237631068216985
138   distance: 6303
139   ratio: 0.49515287099179717
140
141   Parisa deid_Date_Parisa.py : FeverDreams deid.py
142
143   The similarity results tested by Levenshtein are following:
144   jaro: 0.7367499790096219
145   distance: 5355
146   ratio: 0.6577508257079431
147
148   Parisa deid_Date_Parisa.py : FeverDreams deid_phone.py
149
150   The similarity results tested by Levenshtein are following:
151   jaro: 0.7272623539456684
152   distance: 2638
153   ratio: 0.7939667713038818
154
155   Parisa deid_Date_Parisa.py : Itoshikihiro deid-Jie.py
156
157   The similarity results tested by Levenshtein are following:
158   jaro: 0.7322073315167268
159   distance: 2398
160   ratio: 0.8149387693009812
161
162   Parisa deid_Date_Parisa.py : Itoshikihiro deid.py
163
164   The similarity results tested by Levenshtein are following:
165   jaro: 0.7322073315167268
166   distance: 2398
167   ratio: 0.8149387693009812
168
169   Parisa deid_Date_Parisa.py : Itoshikihiro deid-date-Jie.py
```

```
The similarity results tested by Levenshtein are following:
jaro: 0.5837053487892885
distance: 4556
ratio: 0.5521324980691668

Parisa deid_Date_Parisa.py : Itoshikihiro deid-name-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 0.6426224800866115
distance: 4406
ratio: 0.5817552961521834

Parisa deid_Date_Parisa.py : Nasim de-id-location.py

The similarity results tested by Levenshtein are following:
jaro: 0.7612652290570937
distance: 3969
ratio: 0.6266994691182183

Parisa deid_Date_Parisa.py : Nasim deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.7322073315167268
distance: 2398
ratio: 0.8149387693009812

Parisa deid_Date_Parisa.py : Nasim de-id-date.py

The similarity results tested by Levenshtein are following:
jaro: 0.7908737317285306
distance: 2574
ratio: 0.7787909490037149

Parisa deid_Date_Parisa.py : Nasim de-id-age.py

The similarity results tested by Levenshtein are following:
jaro: 0.7140789999533025
distance: 3698
ratio: 0.6609785563273936

Parisa deid_Date_Parisa.py : Andre deid-Andre_Holder.py

The similarity results tested by Levenshtein are following:
jaro: 0.7089230082292618
distance: 2789
ratio: 0.7516325482427177

Parisa deid_Date_Parisa.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7342713019060986
distance: 2394
```

```
ratio: 0.8124478178368122

Parisa deid_Date_Parisa.py : Mahmoud deid_date_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7271487077544735
distance: 2429
ratio: 0.7919453222606012

Parisa deid_Date_Parisa.py : Mahmoud deid_age_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7362675054664228
distance: 2394
ratio: 0.8026986506746626

Parisa deid_Date_Parisa.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.623545294976766
distance: 14776
ratio: 0.43521914052764626

Parisa deid_Date_Parisa.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.7322073315167268
distance: 2398
ratio: 0.8149387693009812

Parisa deid_Age_Parisa.py : Shaojun deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.7243258176235227
distance: 2638
ratio: 0.796168811733014

Parisa deid_Age_Parisa.py : Shaojun deid-Shaojun Yu.py

The similarity results tested by Levenshtein are following:
jaro: 0.7420013817826021
distance: 6248
ratio: 0.501756373937677

Parisa deid_Age_Parisa.py : FeverDreams deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.7389892685200682
distance: 5165
ratio: 0.6595311998287487

Parisa deid_Age_Parisa.py : FeverDreams deid_phone.py
```

```
276  The similarity results tested by Levenshtein are following:
277  jaro: 0.7195949762116066
278  distance: 2877
279  ratio: 0.775116734447959
280
281  Parisa deid_Age_Parisa.py : Itoshikihiro deid-Jie.py
282
283  The similarity results tested by Levenshtein are following:
284  jaro: 0.7243258176235227
285  distance: 2638
286  ratio: 0.796168811733014
287
288  Parisa deid_Age_Parisa.py : Itoshikihiro deid.py
289
290  The similarity results tested by Levenshtein are following:
291  jaro: 0.7243258176235227
292  distance: 2638
293  ratio: 0.796168811733014
294
295  Parisa deid_Age_Parisa.py : Itoshikihiro deid-date-Jie.py
296
297  The similarity results tested by Levenshtein are following:
298  jaro: 0.5723869285370724
299  distance: 4791
300  ratio: 0.535804549283909
301
302  Parisa deid_Age_Parisa.py : Itoshikihiro deid-name-Jie.py
303
304  The similarity results tested by Levenshtein are following:
305  jaro: 0.6343294942401324
306  distance: 4631
307  ratio: 0.5693430656934306
308
309  Parisa deid_Age_Parisa.py : Nasim de-id-location.py
310
311  The similarity results tested by Levenshtein are following:
312  jaro: 0.7590299153664958
313  distance: 3979
314  ratio: 0.6337227861840005
315
316  Parisa deid_Age_Parisa.py : Nasim deid.py
317
318  The similarity results tested by Levenshtein are following:
319  jaro: 0.7243258176235227
320  distance: 2638
321  ratio: 0.796168811733014
322
323  Parisa deid_Age_Parisa.py : Nasim de-id-date.py
324
325  The similarity results tested by Levenshtein are following:
326  jaro: 0.7832572893510522
327  distance: 2708
328  ratio: 0.7639462122220743
```

```
329
330   Parisa deid_Age_Parisa.py : Nasim de-id-age.py
331
332   The similarity results tested by Levenshtein are following:
333   jaro: 0.7116282243114193
334   distance: 3868
335   ratio: 0.657752024366689
336
337   Parisa deid_Age_Parisa.py : Andre deid-Andre_Holder.py
338
339   The similarity results tested by Levenshtein are following:
340   jaro: 0.7016318793130444
341   distance: 3015
342   ratio: 0.7350859453993933
343
344   Parisa deid_Age_Parisa.py : Mahmoud deid_ssn_mahmoud.py
345
346   The similarity results tested by Levenshtein are following:
347   jaro: 0.7272199841071779
348   distance: 2629
349   ratio: 0.7955495818399044
350
351   Parisa deid_Age_Parisa.py : Mahmoud deid_date_mahmoud.py
352
353   The similarity results tested by Levenshtein are following:
354   jaro: 0.7194463696118539
355   distance: 2639
356   ratio: 0.7779224537037037
357
358   Parisa deid_Age_Parisa.py : Mahmoud deid_age_mahmoud.py
359
360   The similarity results tested by Levenshtein are following:
361   jaro: 0.7291634696614113
362   distance: 2600
363   ratio: 0.7920631408128642
364
365   Parisa deid_Age_Parisa.py : ChenLin deid-ChenLin.py
366
367   The similarity results tested by Levenshtein are following:
368   jaro: 0.6245976809757386
369   distance: 14610
370   ratio: 0.44601469756924816
371
372   Parisa deid_Age_Parisa.py : ChenLin deid.py
373
374   The similarity results tested by Levenshtein are following:
375   jaro: 0.7243258176235227
376   distance: 2638
377   ratio: 0.796168811733014
378
379   Parisa deid_Location_Parisa.py : Shaojun deid.py
380
381   The similarity results tested by Levenshtein are following:
```

```
jaro: 0.6437365646226846
distance: 5972
ratio: 0.6351456893962151

Parisa deid_Location_Parisa.py : Shaojun deid-Shaojun Yu.py

The similarity results tested by Levenshtein are following:
jaro: 0.7234994300222922
distance: 7455
ratio: 0.4926663800105107

Parisa deid_Location_Parisa.py : FeverDreams deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.7655451703892462
distance: 6373
ratio: 0.6265762279783311

Parisa deid_Location_Parisa.py : FeverDreams deid_phone.py

The similarity results tested by Levenshtein are following:
jaro: 0.6401688216903018
distance: 6187
ratio: 0.6181532701250075

Parisa deid_Location_Parisa.py : Itoshikihiro deid-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 0.6437365646226846
distance: 5972
ratio: 0.6351456893962151

Parisa deid_Location_Parisa.py : Itoshikihiro deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.6437365646226846
distance: 5972
ratio: 0.6351456893962151

Parisa deid_Location_Parisa.py : Itoshikihiro deid-date-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 0.5295364901509254
distance: 8121
ratio: 0.41594614216883374

Parisa deid_Location_Parisa.py : Itoshikihiro deid-name-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 0.5901775270985143
distance: 7886
ratio: 0.4509061939852619
```

```
435  Parisa deid_Location_Parisa.py : Nasim de-id-location.py
436
437  The similarity results tested by Levenshtein are following:
438  jaro: 0.6984485606975257
439  distance: 6280
440  ratio: 0.579075168918919
441
442  Parisa deid_Location_Parisa.py : Nasim deid.py
443
444  The similarity results tested by Levenshtein are following:
445  jaro: 0.6437365646226846
446  distance: 5972
447  ratio: 0.6351456893962151
448
449  Parisa deid_Location_Parisa.py : Nasim de-id-date.py
450
451  The similarity results tested by Levenshtein are following:
452  jaro: 0.6950032392233699
453  distance: 5486
454  ratio: 0.6526798885428619
455
456  Parisa deid_Location_Parisa.py : Nasim de-id-age.py
457
458  The similarity results tested by Levenshtein are following:
459  jaro: 0.6430648750237822
460  distance: 6834
461  ratio: 0.5479632063074902
462
463  Parisa deid_Location_Parisa.py : Andre deid-Andre_Holder.py
464
465  The similarity results tested by Levenshtein are following:
466  jaro: 0.6186117818786692
467  distance: 6213
468  ratio: 0.5934489402697495
469
470  Parisa deid_Location_Parisa.py : Mahmoud deid_ssn_mahmoud.py
471
472  The similarity results tested by Levenshtein are following:
473  jaro: 0.647447643781852
474  distance: 5960
475  ratio: 0.6351586397169076
476
477  Parisa deid_Location_Parisa.py : Mahmoud deid_date_mahmoud.py
478
479  The similarity results tested by Levenshtein are following:
480  jaro: 0.6349372362754272
481  distance: 5903
482  ratio: 0.625431160479392
483
484  Parisa deid_Location_Parisa.py : Mahmoud deid_age_mahmoud.py
485
486  The similarity results tested by Levenshtein are following:
487  jaro: 0.6470411874713099
```

```
488    distance: 5921
489    ratio: 0.633566919396603
490
491    Parisa deid_Location_Parisa.py : ChenLin deid-ChenLin.py
492
493    The similarity results tested by Levenshtein are following:
494    jaro: 0.6639069664685247
495    distance: 13990
496    ratio: 0.46123159727718854
497
498    Parisa deid_Location_Parisa.py : ChenLin deid.py
499
500    The similarity results tested by Levenshtein are following:
501    jaro: 0.6437365646226846
502    distance: 5972
503    ratio: 0.6351456893962151
504
505    Shaojun deid.py : FeverDreams deid.py
506
507    The similarity results tested by Levenshtein are following:
508    jaro: 0.6684480457968126
509    distance: 5400
510    ratio: 0.6632829908393166
511
512    Shaojun deid.py : FeverDreams deid_phone.py
513
514    The similarity results tested by Levenshtein are following:
515    jaro: 0.822566622827735
516    distance: 261
517    ratio: 0.9741347227391143
518
519    Shaojun deid.py : Itoshikihiro deid-Jie.py
520
521    The similarity results tested by Levenshtein are following:
522    jaro: 1.0
523    distance: 0
524    ratio: 1.0
525
526    Shaojun deid.py : Itoshikihiro deid.py
527
528    The similarity results tested by Levenshtein are following:
529    jaro: 1.0
530    distance: 0
531    ratio: 1.0
532
533    Shaojun deid.py : Itoshikihiro deid-date-Jie.py
534
535    The similarity results tested by Levenshtein are following:
536    jaro: 0.6557711323609124
537    distance: 2317
538    ratio: 0.6839400428265524
539
540    Shaojun deid.py : Itoshikihiro deid-name-Jie.py
```

```
The similarity results tested by Levenshtein are following:
jaro: 0.7251300664288114
distance: 2450
ratio: 0.6850410722006053

Shaojun deid.py : Nasim de-id-location.py

The similarity results tested by Levenshtein are following:
jaro: 0.7288723120536037
distance: 3611
ratio: 0.6504225995583645

Shaojun deid.py : Nasim deid.py

The similarity results tested by Levenshtein are following:
jaro: 1.0
distance: 0
ratio: 1.0

Shaojun deid.py : Nasim de-id-date.py

The similarity results tested by Levenshtein are following:
jaro: 0.7532462007044755
distance: 2080
ratio: 0.8198847262247838

Shaojun deid.py : Nasim de-id-age.py

The similarity results tested by Levenshtein are following:
jaro: 0.779738456966431
distance: 2053
ratio: 0.7371695178849145

Shaojun deid.py : Andre deid-Andre_Holder.py

The similarity results tested by Levenshtein are following:
jaro: 0.7873682872182086
distance: 1287
ratio: 0.8806998939554613

Shaojun deid.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.8231258743340041
distance: 211
ratio: 0.9756950837783097

Shaojun deid.py : Mahmoud deid_date_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7991964384755024
distance: 702
```

```
ratio: 0.9272179918540818

Shaojun deid.py : Mahmoud deid_age_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.8142641234970611
distance: 427
ratio: 0.950575859254557

Shaojun deid.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.5947612031278235
distance: 14970
ratio: 0.41918212151780865

Shaojun deid.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 1.0
distance: 0
ratio: 1.0

Shaojun deid-Shaojun Yu.py : FeverDreams deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.7124344572682443
distance: 7468
ratio: 0.4725565013208101

Shaojun deid-Shaojun Yu.py : FeverDreams deid_phone.py

The similarity results tested by Levenshtein are following:
jaro: 0.6567340587947946
distance: 6251
ratio: 0.4920846082213649

Shaojun deid-Shaojun Yu.py : Itoshikihiro deid-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 0.6622267258095486
distance: 6217
ratio: 0.49563492063492065

Shaojun deid-Shaojun Yu.py : Itoshikihiro deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.6622267258095486
distance: 6217
ratio: 0.49563492063492065

Shaojun deid-Shaojun Yu.py : Itoshikihiro deid-date-Jie.py
```

```
The similarity results tested by Levenshtein are following:
jaro: 0.5727770920659783
distance: 6961
ratio: 0.41611624834874505

Shaojun deid-Shaojun Yu.py : Itoshikihiro deid-name-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 0.6137336138928172
distance: 6695
ratio: 0.45797015807357067

Shaojun deid-Shaojun Yu.py : Nasim de-id-location.py

The similarity results tested by Levenshtein are following:
jaro: 0.7306528551812068
distance: 6176
ratio: 0.49520638383374477

Shaojun deid-Shaojun Yu.py : Nasim deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.6622267258095486
distance: 6217
ratio: 0.49563492063492065

Shaojun deid-Shaojun Yu.py : Nasim de-id-date.py

The similarity results tested by Levenshtein are following:
jaro: 0.7248433566644152
distance: 6140
ratio: 0.5056621766599118

Shaojun deid-Shaojun Yu.py : Nasim de-id-age.py

The similarity results tested by Levenshtein are following:
jaro: 0.676308203734497
distance: 6059
ratio: 0.5121903134652034

Shaojun deid-Shaojun Yu.py : Andre deid-Andre_Holder.py

The similarity results tested by Levenshtein are following:
jaro: 0.660990082325941
distance: 6312
ratio: 0.47660556338930904

Shaojun deid-Shaojun Yu.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.6646610308045853
distance: 6233
ratio: 0.49471877475574333
```

```
Shaojun deid-Shaojun Yu.py : Mahmoud deid_date_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.6736615772550122
distance: 6318
ratio: 0.48292682926829267


Shaojun deid-Shaojun Yu.py : Mahmoud deid_age_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.6694987258141047
distance: 6264
ratio: 0.49030235747404166


Shaojun deid-Shaojun Yu.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.6073715186342309
distance: 14620
ratio: 0.4179640718562874


Shaojun deid-Shaojun Yu.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.6622267258095486
distance: 6217
ratio: 0.49563492063492065


FeverDreams deid.py : Itoshikihiro deid-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 0.6684480457968126
distance: 5400
ratio: 0.6632829908393166


FeverDreams deid.py : Itoshikihiro deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.6684480457968126
distance: 5400
ratio: 0.6632829908393166


FeverDreams deid.py : Itoshikihiro deid-date-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 0.609015309758357
distance: 7108
ratio: 0.5066157413722548


FeverDreams deid.py : Itoshikihiro deid-name-Jie.py

The similarity results tested by Levenshtein are following:
```

```
jaro: 0.605252252532336
distance: 7667
ratio: 0.44119665157129134

FeverDreams deid.py : Nasim de-id-location.py

The similarity results tested by Levenshtein are following:
jaro: 0.7000749155711296
distance: 6385
ratio: 0.5432674072067191

FeverDreams deid.py : Nasim deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.6684480457968126
distance: 5400
ratio: 0.6632829908393166

FeverDreams deid.py : Nasim de-id-date.py

The similarity results tested by Levenshtein are following:
jaro: 0.7144974163880845
distance: 5406
ratio: 0.6406197372852812

FeverDreams deid.py : Nasim de-id-age.py

The similarity results tested by Levenshtein are following:
jaro: 0.6466321019112898
distance: 6771
ratio: 0.5293791915338707

FeverDreams deid.py : Andre deid-Andre_Holder.py

The similarity results tested by Levenshtein are following:
jaro: 0.6590157499549273
distance: 6048
ratio: 0.6147373482389711

FeverDreams deid.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.6691917126632388
distance: 5475
ratio: 0.6543499752842313

FeverDreams deid.py : Mahmoud deid_date_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.6729055762370497
distance: 5565
ratio: 0.6462445835339432
```

```
FeverDreams deid.py : Mahmoud deid_age_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.6728284776142273
distance: 5506
ratio: 0.6462780598201725

FeverDreams deid.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.6602505669234963
distance: 11383
ratio: 0.6153202984306663

FeverDreams deid.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.6684480457968126
distance: 5400
ratio: 0.6632829908393166

FeverDreams deid_phone.py : Itoshikihiro deid-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 0.822566622827735
distance: 261
ratio: 0.9741347227391143

FeverDreams deid_phone.py : Itoshikihiro deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.822566622827735
distance: 261
ratio: 0.9741347227391143

FeverDreams deid_phone.py : Itoshikihiro deid-date-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 0.6593279276576074
distance: 2496
ratio: 0.6576615517614005

FeverDreams deid_phone.py : Itoshikihiro deid-name-Jie.py

The similarity results tested by Levenshtein are following:
jaro: 0.72880103895691
distance: 2485
ratio: 0.6757582369626882

FeverDreams deid_phone.py : Nasim de-id-location.py

The similarity results tested by Levenshtein are following:
jaro: 0.7244322067261076
```

```
distance: 3781
ratio: 0.6330957308193454

FeverDreams deid_phone.py : Nasim deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.822566622827735
distance: 261
ratio: 0.9741347227391143

FeverDreams deid_phone.py : Nasim de-id-date.py

The similarity results tested by Levenshtein are following:
jaro: 0.7477074996913037
distance: 2270
ratio: 0.8025149121392875

FeverDreams deid_phone.py : Nasim de-id-age.py

The similarity results tested by Levenshtein are following:
jaro: 0.7763666382428639
distance: 2247
ratio: 0.7131396957123098

FeverDreams deid_phone.py : Andre deid-Andre_Holder.py

The similarity results tested by Levenshtein are following:
jaro: 0.7794423204248448
distance: 1509
ratio: 0.8577025823686554

FeverDreams deid_phone.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.8088390049312494
distance: 468
ratio: 0.9497030438010393

FeverDreams deid_phone.py : Mahmoud deid_date_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7865465859811686
distance: 918
ratio: 0.9057815845824411

FeverDreams deid_phone.py : Mahmoud deid_age_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.8010374003352917
distance: 683
ratio: 0.9251439539347409

FeverDreams deid_phone.py : ChenLin deid-ChenLin.py
```

```
The similarity results tested by Levenshtein are following:
jaro: 0.5867514442377398
distance: 15168
ratio: 0.4089068825910931

FeverDreams deid_phone.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.822566622827735
distance: 261
ratio: 0.9741347227391143

Itoshikihiro deid-Jie.py : Nasim de-id-location.py

The similarity results tested by Levenshtein are following:
jaro: 0.7288723120536037
distance: 3611
ratio: 0.6504225995583645

Itoshikihiro deid-Jie.py : Nasim deid.py

The similarity results tested by Levenshtein are following:
jaro: 1.0
distance: 0
ratio: 1.0

Itoshikihiro deid-Jie.py : Nasim de-id-date.py

The similarity results tested by Levenshtein are following:
jaro: 0.7532462007044755
distance: 2080
ratio: 0.8198847262247838

Itoshikihiro deid-Jie.py : Nasim de-id-age.py

The similarity results tested by Levenshtein are following:
jaro: 0.779738456966431
distance: 2053
ratio: 0.7371695178849145

Itoshikihiro deid-Jie.py : Andre deid-Andre_Holder.py

The similarity results tested by Levenshtein are following:
jaro: 0.7873682872182086
distance: 1287
ratio: 0.8806998939554613

Itoshikihiro deid-Jie.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.8231258743340041
distance: 211
```

```
ratio: 0.9756950837783097

Itoshikihiro deid-Jie.py : Mahmoud deid_date_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7991964384755024
distance: 702
ratio: 0.9272179918540818

Itoshikihiro deid-Jie.py : Mahmoud deid_age_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.8142641234970611
distance: 427
ratio: 0.950575859254557

Itoshikihiro deid-Jie.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.5947612031278235
distance: 14970
ratio: 0.41918212151780865

Itoshikihiro deid-Jie.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 1.0
distance: 0
ratio: 1.0

Itoshikihiro deid.py : Nasim de-id-location.py

The similarity results tested by Levenshtein are following:
jaro: 0.7288723120536037
distance: 3611
ratio: 0.6504225995583645

Itoshikihiro deid.py : Nasim deid.py

The similarity results tested by Levenshtein are following:
jaro: 1.0
distance: 0
ratio: 1.0

Itoshikihiro deid.py : Nasim de-id-date.py

The similarity results tested by Levenshtein are following:
jaro: 0.7532462007044755
distance: 2080
ratio: 0.8198847262247838

Itoshikihiro deid.py : Nasim de-id-age.py
```

```
The similarity results tested by Levenshtein are following:
jaro: 0.779738456966431
distance: 2053
ratio: 0.7371695178849145

Itoshikihiro deid.py : Andre deid-Andre_Holder.py

The similarity results tested by Levenshtein are following:
jaro: 0.7873682872182086
distance: 1287
ratio: 0.8806998939554613

Itoshikihiro deid.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.8231258743340041
distance: 211
ratio: 0.9756950837783097

Itoshikihiro deid.py : Mahmoud deid_date_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7991964384755024
distance: 702
ratio: 0.9272179918540818

Itoshikihiro deid.py : Mahmoud deid_age_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.8142641234970611
distance: 427
ratio: 0.950575859254557

Itoshikihiro deid.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.5947612031278235
distance: 14970
ratio: 0.41918212151780865

Itoshikihiro deid.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 1.0
distance: 0
ratio: 1.0

Itoshikihiro deid-date-Jie.py : Nasim de-id-location.py

The similarity results tested by Levenshtein are following:
jaro: 0.5834987388663261
distance: 5070
ratio: 0.4723773520061861
```

```
Itoshikihiro deid-date-Jie.py : Nasim deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.6557711323609124
distance: 2317
ratio: 0.6839400428265524

Itoshikihiro deid-date-Jie.py : Nasim de-id-date.py

The similarity results tested by Levenshtein are following:
jaro: 0.599676605908865
distance: 4092
ratio: 0.5566466630296417

Itoshikihiro deid-date-Jie.py : Nasim de-id-age.py

The similarity results tested by Levenshtein are following:
jaro: 0.6548455298038197
distance: 3123
ratio: 0.5599237045671294

Itoshikihiro deid-date-Jie.py : Andre deid-Andre_Holder.py

The similarity results tested by Levenshtein are following:
jaro: 0.6721955902223019
distance: 2696
ratio: 0.6760333944206882

Itoshikihiro deid-date-Jie.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.6516199526483168
distance: 2361
ratio: 0.6767719897523484

Itoshikihiro deid-date-Jie.py : Mahmoud deid_date_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.6731968568452693
distance: 2709
ratio: 0.6712244897959184

Itoshikihiro deid-date-Jie.py : Mahmoud deid_age_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.6593629096943436
distance: 2477
ratio: 0.6734501206335886

Itoshikihiro deid-date-Jie.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
```

```
jaro: 0.5057462566712471
distance: 17175
ratio: 0.264168039538715

Itoshikihiro deid-date-Jie.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.6557711323609124
distance: 2317
ratio: 0.6839400428265524

Itoshikihiro deid-name-Jie.py : Nasim de-id-location.py

The similarity results tested by Levenshtein are following:
jaro: 0.6466414665495342
distance: 4919
ratio: 0.5007358670245

Itoshikihiro deid-name-Jie.py : Nasim deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.7251300664288114
distance: 2450
ratio: 0.6850410722006053

Itoshikihiro deid-name-Jie.py : Nasim de-id-date.py

The similarity results tested by Levenshtein are following:
jaro: 0.6621100984577832
distance: 4014
ratio: 0.5767186067827681

Itoshikihiro deid-name-Jie.py : Nasim de-id-age.py

The similarity results tested by Levenshtein are following:
jaro: 0.7237136582012825
distance: 2995
ratio: 0.5820943416408172

Itoshikihiro deid-name-Jie.py : Andre deid-Andre_Holder.py

The similarity results tested by Levenshtein are following:
jaro: 0.7007765437113408
distance: 2942
ratio: 0.6420793096363263

Itoshikihiro deid-name-Jie.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.723291008331523
distance: 2484
ratio: 0.6818965517241379
```

```
1177   Itoshikihiro deid-name-Jie.py : Mahmoud deid_date_mahmoud.py
1178
1179   The similarity results tested by Levenshtein are following:
1180   jaro: 0.703185806363572
1181   distance: 2923
1182   ratio: 0.6497116968698518
1183
1184   Itoshikihiro deid-name-Jie.py : Mahmoud deid_age_mahmoud.py
1185
1186   The similarity results tested by Levenshtein are following:
1187   jaro: 0.7195861618718049
1188   distance: 2624
1189   ratio: 0.6727368978295395
1190
1191   Itoshikihiro deid-name-Jie.py : ChenLin deid-ChenLin.py
1192
1193   The similarity results tested by Levenshtein are following:
1194   jaro: 0.5551738091376321
1195   distance: 16838
1196   ratio: 0.29224537037037035
1197
1198   Itoshikihiro deid-name-Jie.py : ChenLin deid.py
1199
1200   The similarity results tested by Levenshtein are following:
1201   jaro: 0.7251300664288114
1202   distance: 2450
1203   ratio: 0.6850410722006053
1204
1205   Nasim de-id-location.py : Andre deid-Andre_Holder.py
1206
1207   The similarity results tested by Levenshtein are following:
1208   jaro: 0.7047714451423163
1209   distance: 3828
1210   ratio: 0.6178479618068307
1211
1212   Nasim de-id-location.py : Mahmoud deid_ssn_mahmoud.py
1213
1214   The similarity results tested by Levenshtein are following:
1215   jaro: 0.7314117229024557
1216   distance: 3620
1217   ratio: 0.647215257199301
1218
1219   Nasim de-id-location.py : Mahmoud deid_date_mahmoud.py
1220
1221   The similarity results tested by Levenshtein are following:
1222   jaro: 0.7220739788810301
1223   distance: 3710
1224   ratio: 0.6282645479290812
1225
1226   Nasim de-id-location.py : Mahmoud deid_age_mahmoud.py
1227
1228   The similarity results tested by Levenshtein are following:
1229   jaro: 0.7325388852011249
```

```
distance: 3607
ratio: 0.6431037070388714

Nasim de-id-location.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.5964311285440955
distance: 15296
ratio: 0.38221778933494815

Nasim de-id-location.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.7288723120536037
distance: 3611
ratio: 0.6504225995583645

Nasim deid.py : Andre deid-Andre_Holder.py

The similarity results tested by Levenshtein are following:
jaro: 0.7873682872182086
distance: 1287
ratio: 0.8806998939554613

Nasim deid.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.8231258743340041
distance: 211
ratio: 0.9756950837783097

Nasim deid.py : Mahmoud deid_date_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7991964384755024
distance: 702
ratio: 0.9272179918540818

Nasim deid.py : Mahmoud deid_age_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.8142641234970611
distance: 427
ratio: 0.950575859254557

Nasim deid.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.5947612031278235
distance: 14970
ratio: 0.41918212151780865

Nasim deid.py : ChenLin deid.py
```

```
The similarity results tested by Levenshtein are following:
jaro: 1.0
distance: 0
ratio: 1.0

Nasim de-id-date.py : Andre deid-Andre_Holder.py

The similarity results tested by Levenshtein are following:
jaro: 0.7324080950103764
distance: 2606
ratio: 0.7599815014644674

Nasim de-id-date.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.755697263461831
distance: 2055
ratio: 0.818849840255591

Nasim de-id-date.py : Mahmoud deid_date_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7502499723558853
distance: 2183
ratio: 0.7896849907350216

Nasim de-id-date.py : Mahmoud deid_age_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7573211946021731
distance: 2103
ratio: 0.8048876625936146

Nasim de-id-date.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.6095559054011043
distance: 15158
ratio: 0.4195829687955672

Nasim de-id-date.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.7532462007044755
distance: 2080
ratio: 0.8198847262247838

Nasim de-id-age.py : Andre deid-Andre_Holder.py

The similarity results tested by Levenshtein are following:
jaro: 0.7644899028373094
distance: 2576
```

```
ratio: 0.6962236046613511

Nasim de-id-age.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7806685253695465
distance: 2126
ratio: 0.7290811205401952

Nasim de-id-age.py : Mahmoud deid_date_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.768758933176998
distance: 2479
ratio: 0.7000263365815117

Nasim de-id-age.py : Mahmoud deid_age_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7819688398610026
distance: 2250
ratio: 0.7209636821287306

Nasim de-id-age.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.5679697031819618
distance: 16013
ratio: 0.3486529318541997

Nasim de-id-age.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.779738456966431
distance: 2053
ratio: 0.7371695178849145

Andre deid-Andre_Holder.py : Mahmoud deid_ssn_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7840895490604356
distance: 1252
ratio: 0.8802891396332864

Andre deid-Andre_Holder.py : Mahmoud deid_date_mahmoud.py

The similarity results tested by Levenshtein are following:
jaro: 0.7980585540212525
distance: 1276
ratio: 0.8712635869565217

Andre deid-Andre_Holder.py : Mahmoud deid_age_mahmoud.py
```

```
The similarity results tested by Levenshtein are following:
jaro: 0.7872139751164751
distance: 1260
ratio: 0.8707967677469807

Andre deid-Andre_Holder.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.5616178186306954
distance: 15731
ratio: 0.38033211456429006

Andre deid-Andre_Holder.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.7873682872182086
distance: 1287
ratio: 0.8806998939554613

Mahmoud deid_ssn_mahmoud.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.5888555109093102
distance: 15171
ratio: 0.41027827300209285

Mahmoud deid_ssn_mahmoud.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.8231258743340041
distance: 211
ratio: 0.9756950837783097

Mahmoud deid_date_mahmoud.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.5740885489860531
distance: 15654
ratio: 0.399252877944652

Mahmoud deid_date_mahmoud.py : ChenLin deid.py

The similarity results tested by Levenshtein are following:
jaro: 0.7991964384755024
distance: 702
ratio: 0.9272179918540818

Mahmoud deid_age_mahmoud.py : ChenLin deid-ChenLin.py

The similarity results tested by Levenshtein are following:
jaro: 0.5867488786590064
distance: 15387
ratio: 0.4032040667000424
```

```
1442
1443   Mahmoud deid_age_mahmoud.py : ChenLin deid.py
1444
1445   The similarity results tested by Levenshtein are following:
1446   jaro: 0.8142641234970611
1447   distance: 427
1448   ratio: 0.950575859254557
```

Tensorflow Virtualization Similarity Score

Semantic Textual Similarity
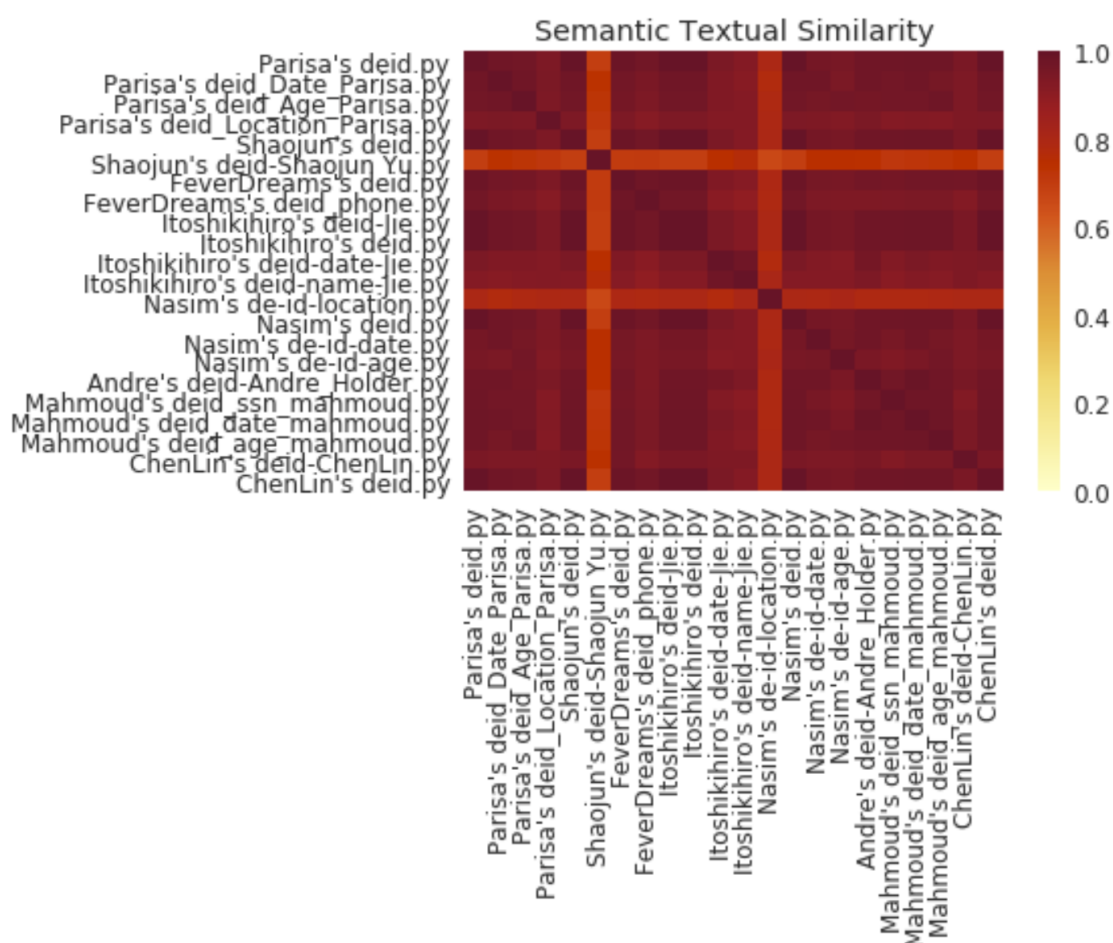
STS Benchmark

```
1    Parisa's deid.py compares to Parisa's deid_Date_Parisa.py
2        STS Benchmark Score is 0.77391994
3    Parisa's deid.py compares to Parisa's deid_Age_Parisa.py
4        STS Benchmark Score is 0.7577548
5    Parisa's deid.py compares to Parisa's deid_Location_Parisa.py
6        STS Benchmark Score is 0.6646246
7    Parisa's deid.py compares to Shaojun's deid.py
8        STS Benchmark Score is 0.9995117
9    Parisa's deid.py compares to Shaojun's deid-Shaojun Yu.py
10       STS Benchmark Score is 0.20311517
11   Parisa's deid.py compares to FeverDreams's deid.py
12       STS Benchmark Score is 0.8443497
```

```
13  Parisa's deid.py compares to FeverDreams's deid_phone.py
14      STS Benchmark Score is 0.75724745
15  Parisa's deid.py compares to Itoshikihiro's deid-Jie.py
16      STS Benchmark Score is 0.9995117
17  Parisa's deid.py compares to Itoshikihiro's deid.py
18      STS Benchmark Score is 0.9995117
19  Parisa's deid.py compares to Itoshikihiro's deid-date-Jie.py
20      STS Benchmark Score is 0.6723099
21  Parisa's deid.py compares to Itoshikihiro's deid-name-Jie.py
22      STS Benchmark Score is 0.61155987
23  Parisa's deid.py compares to Nasim's de-id-location.py
24      STS Benchmark Score is 0.35966337
25  Parisa's deid.py compares to Nasim's deid.py
26      STS Benchmark Score is 0.9995117
27  Parisa's deid.py compares to Nasim's de-id-date.py
28      STS Benchmark Score is 0.7326237
29  Parisa's deid.py compares to Nasim's de-id-age.py
30      STS Benchmark Score is 0.6962836
31  Parisa's deid.py compares to Andre's deid-Andre_Holder.py
32      STS Benchmark Score is 0.7823216
33  Parisa's deid.py compares to Mahmoud's deid_ssn_mahmoud.py
34      STS Benchmark Score is 0.7982335
35  Parisa's deid.py compares to Mahmoud's deid_date_mahmoud.py
36      STS Benchmark Score is 0.8076906
37  Parisa's deid.py compares to Mahmoud's deid_age_mahmoud.py
38      STS Benchmark Score is 0.8005162
39  Parisa's deid.py compares to ChenLin's deid-ChenLin.py
40      STS Benchmark Score is 0.6855277
41  Parisa's deid.py compares to ChenLin's deid.py
42      STS Benchmark Score is 0.9995117
43  Parisa's deid_Date_Parisa.py compares to Parisa's deid_Age_Parisa.py
44      STS Benchmark Score is 0.7796645
45  Parisa's deid_Date_Parisa.py compares to Parisa's deid_Location_Parisa.py
46      STS Benchmark Score is 0.6748899
47  Parisa's deid_Date_Parisa.py compares to Shaojun's deid.py
48      STS Benchmark Score is 0.77391994
49  Parisa's deid_Date_Parisa.py compares to Shaojun's deid-Shaojun Yu.py
50      STS Benchmark Score is 0.2562076
51  Parisa's deid_Date_Parisa.py compares to FeverDreams's deid.py
52      STS Benchmark Score is 0.7462578
53  Parisa's deid_Date_Parisa.py compares to FeverDreams's deid_phone.py
54      STS Benchmark Score is 0.6814806
55  Parisa's deid_Date_Parisa.py compares to Itoshikihiro's deid-Jie.py
56      STS Benchmark Score is 0.77391994
57  Parisa's deid_Date_Parisa.py compares to Itoshikihiro's deid.py
58      STS Benchmark Score is 0.77391994
59  Parisa's deid_Date_Parisa.py compares to Itoshikihiro's deid-date-Jie.py
60      STS Benchmark Score is 0.64279383
61  Parisa's deid_Date_Parisa.py compares to Itoshikihiro's deid-name-Jie.py
62      STS Benchmark Score is 0.5898832
63  Parisa's deid_Date_Parisa.py compares to Nasim's de-id-location.py
64      STS Benchmark Score is 0.32214737
65  Parisa's deid_Date_Parisa.py compares to Nasim's deid.py
```

```
    66        STS Benchmark Score is 0.77391994
    67   Parisa's deid_Date_Parisa.py compares to Nasim's de-id-date.py
    68        STS Benchmark Score is 0.7804253
    69   Parisa's deid_Date_Parisa.py compares to Nasim's de-id-age.py
    70        STS Benchmark Score is 0.66802084
    71   Parisa's deid_Date_Parisa.py compares to Andre's deid-Andre_Holder.py
    72        STS Benchmark Score is 0.7740681
    73   Parisa's deid_Date_Parisa.py compares to Mahmoud's deid_ssn_mahmoud.py
    74        STS Benchmark Score is 0.742605
    75   Parisa's deid_Date_Parisa.py compares to Mahmoud's deid_date_mahmoud.py
    76        STS Benchmark Score is 0.77550113
    77   Parisa's deid_Date_Parisa.py compares to Mahmoud's deid_age_mahmoud.py
    78        STS Benchmark Score is 0.71769524
    79   Parisa's deid_Date_Parisa.py compares to ChenLin's deid-ChenLin.py
    80        STS Benchmark Score is 0.6668976
    81   Parisa's deid_Date_Parisa.py compares to ChenLin's deid.py
    82        STS Benchmark Score is 0.77391994
    83   Parisa's deid_Age_Parisa.py compares to Parisa's deid_Location_Parisa.py
    84        STS Benchmark Score is 0.66112536
    85   Parisa's deid_Age_Parisa.py compares to Shaojun's deid.py
    86        STS Benchmark Score is 0.7577548
    87   Parisa's deid_Age_Parisa.py compares to Shaojun's deid-Shaojun Yu.py
    88        STS Benchmark Score is 0.24589944
    89   Parisa's deid_Age_Parisa.py compares to FeverDreams's deid.py
    90        STS Benchmark Score is 0.73141736
    91   Parisa's deid_Age_Parisa.py compares to FeverDreams's deid_phone.py
    92        STS Benchmark Score is 0.66182876
    93   Parisa's deid_Age_Parisa.py compares to Itoshikihiro's deid-Jie.py
    94        STS Benchmark Score is 0.7577548
    95   Parisa's deid_Age_Parisa.py compares to Itoshikihiro's deid.py
    96        STS Benchmark Score is 0.7577548
    97   Parisa's deid_Age_Parisa.py compares to Itoshikihiro's deid-date-Jie.py
    98        STS Benchmark Score is 0.6414424
    99   Parisa's deid_Age_Parisa.py compares to Itoshikihiro's deid-name-Jie.py
   100        STS Benchmark Score is 0.604254
   101   Parisa's deid_Age_Parisa.py compares to Nasim's de-id-location.py
   102        STS Benchmark Score is 0.3404339
   103   Parisa's deid_Age_Parisa.py compares to Nasim's deid.py
   104        STS Benchmark Score is 0.7577548
   105   Parisa's deid_Age_Parisa.py compares to Nasim's de-id-date.py
   106        STS Benchmark Score is 0.708049
   107   Parisa's deid_Age_Parisa.py compares to Nasim's de-id-age.py
   108        STS Benchmark Score is 0.7354356
   109   Parisa's deid_Age_Parisa.py compares to Andre's deid-Andre_Holder.py
   110        STS Benchmark Score is 0.7289948
   111   Parisa's deid_Age_Parisa.py compares to Mahmoud's deid_ssn_mahmoud.py
   112        STS Benchmark Score is 0.7395792
   113   Parisa's deid_Age_Parisa.py compares to Mahmoud's deid_date_mahmoud.py
   114        STS Benchmark Score is 0.73591864
   115   Parisa's deid_Age_Parisa.py compares to Mahmoud's deid_age_mahmoud.py
   116        STS Benchmark Score is 0.78681964
   117   Parisa's deid_Age_Parisa.py compares to ChenLin's deid-ChenLin.py
   118        STS Benchmark Score is 0.66384745
```

```
119   Parisa's deid_Age_Parisa.py compares to ChenLin's deid.py
120       STS Benchmark Score is 0.7577548
121   Parisa's deid_Location_Parisa.py compares to Shaojun's deid.py
122       STS Benchmark Score is 0.6646246
123   Parisa's deid_Location_Parisa.py compares to Shaojun's deid-Shaojun Yu.py
124       STS Benchmark Score is 0.23053527
125   Parisa's deid_Location_Parisa.py compares to FeverDreams's deid.py
126       STS Benchmark Score is 0.66578555
127   Parisa's deid_Location_Parisa.py compares to FeverDreams's deid_phone.py
128       STS Benchmark Score is 0.5935304
129   Parisa's deid_Location_Parisa.py compares to Itoshikihiro's deid-Jie.py
130       STS Benchmark Score is 0.6646246
131   Parisa's deid_Location_Parisa.py compares to Itoshikihiro's deid.py
132       STS Benchmark Score is 0.6646246
133   Parisa's deid_Location_Parisa.py compares to Itoshikihiro's deid-date-Jie.py
134       STS Benchmark Score is 0.6217836
135   Parisa's deid_Location_Parisa.py compares to Itoshikihiro's deid-name-Jie.py
136       STS Benchmark Score is 0.6199389
137   Parisa's deid_Location_Parisa.py compares to Nasim's de-id-location.py
138       STS Benchmark Score is 0.35741293
139   Parisa's deid_Location_Parisa.py compares to Nasim's deid.py
140       STS Benchmark Score is 0.6646246
141   Parisa's deid_Location_Parisa.py compares to Nasim's de-id-date.py
142       STS Benchmark Score is 0.64786905
143   Parisa's deid_Location_Parisa.py compares to Nasim's de-id-age.py
144       STS Benchmark Score is 0.62605846
145   Parisa's deid_Location_Parisa.py compares to Andre's deid-Andre_Holder.py
146       STS Benchmark Score is 0.6685132
147   Parisa's deid_Location_Parisa.py compares to Mahmoud's deid_ssn_mahmoud.py
148       STS Benchmark Score is 0.60455835
149   Parisa's deid_Location_Parisa.py compares to Mahmoud's deid_date_mahmoud.py
150       STS Benchmark Score is 0.6281051
151   Parisa's deid_Location_Parisa.py compares to Mahmoud's deid_age_mahmoud.py
152       STS Benchmark Score is 0.6040533
153   Parisa's deid_Location_Parisa.py compares to ChenLin's deid-ChenLin.py
154       STS Benchmark Score is 0.6491078
155   Parisa's deid_Location_Parisa.py compares to ChenLin's deid.py
156       STS Benchmark Score is 0.6646246
157   Shaojun's deid.py compares to Shaojun's deid-Shaojun Yu.py
158       STS Benchmark Score is 0.20311517
159   Shaojun's deid.py compares to FeverDreams's deid.py
160       STS Benchmark Score is 0.8443497
161   Shaojun's deid.py compares to FeverDreams's deid_phone.py
162       STS Benchmark Score is 0.75724745
163   Shaojun's deid.py compares to Itoshikihiro's deid-Jie.py
164       STS Benchmark Score is 0.9995117
165   Shaojun's deid.py compares to Itoshikihiro's deid.py
166       STS Benchmark Score is 0.9995117
167   Shaojun's deid.py compares to Itoshikihiro's deid-date-Jie.py
168       STS Benchmark Score is 0.6723099
169   Shaojun's deid.py compares to Itoshikihiro's deid-name-Jie.py
170       STS Benchmark Score is 0.61155987
171   Shaojun's deid.py compares to Nasim's de-id-location.py
```

```
172        STS Benchmark Score is 0.35966337
173   Shaojun's deid.py compares to Nasim's deid.py
174        STS Benchmark Score is 0.9995117
175   Shaojun's deid.py compares to Nasim's de-id-date.py
176        STS Benchmark Score is 0.7326237
177   Shaojun's deid.py compares to Nasim's de-id-age.py
178        STS Benchmark Score is 0.6962836
179   Shaojun's deid.py compares to Andre's deid-Andre_Holder.py
180        STS Benchmark Score is 0.7823216
181   Shaojun's deid.py compares to Mahmoud's deid_ssn_mahmoud.py
182        STS Benchmark Score is 0.7982335
183   Shaojun's deid.py compares to Mahmoud's deid_date_mahmoud.py
184        STS Benchmark Score is 0.8076906
185   Shaojun's deid.py compares to Mahmoud's deid_age_mahmoud.py
186        STS Benchmark Score is 0.8005162
187   Shaojun's deid.py compares to ChenLin's deid-ChenLin.py
188        STS Benchmark Score is 0.6855277
189   Shaojun's deid.py compares to ChenLin's deid.py
190        STS Benchmark Score is 0.9995117
191   Shaojun's deid-Shaojun Yu.py compares to FeverDreams's deid.py
192        STS Benchmark Score is 0.21289134
193   Shaojun's deid-Shaojun Yu.py compares to FeverDreams's deid_phone.py
194        STS Benchmark Score is 0.21530718
195   Shaojun's deid-Shaojun Yu.py compares to Itoshikihiro's deid-Jie.py
196        STS Benchmark Score is 0.20311517
197   Shaojun's deid-Shaojun Yu.py compares to Itoshikihiro's deid.py
198        STS Benchmark Score is 0.20311517
199   Shaojun's deid-Shaojun Yu.py compares to Itoshikihiro's deid-date-Jie.py
200        STS Benchmark Score is 0.27662957
201   Shaojun's deid-Shaojun Yu.py compares to Itoshikihiro's deid-name-Jie.py
202        STS Benchmark Score is 0.30423874
203   Shaojun's deid-Shaojun Yu.py compares to Nasim's de-id-location.py
204        STS Benchmark Score is 0.1564073
205   Shaojun's deid-Shaojun Yu.py compares to Nasim's deid.py
206        STS Benchmark Score is 0.20311517
207   Shaojun's deid-Shaojun Yu.py compares to Nasim's de-id-date.py
208        STS Benchmark Score is 0.27690572
209   Shaojun's deid-Shaojun Yu.py compares to Nasim's de-id-age.py
210        STS Benchmark Score is 0.27427983
211   Shaojun's deid-Shaojun Yu.py compares to Andre's deid-Andre_Holder.py
212        STS Benchmark Score is 0.26762974
213   Shaojun's deid-Shaojun Yu.py compares to Mahmoud's deid_ssn_mahmoud.py
214        STS Benchmark Score is 0.23090112
215   Shaojun's deid-Shaojun Yu.py compares to Mahmoud's deid_date_mahmoud.py
216        STS Benchmark Score is 0.24041915
217   Shaojun's deid-Shaojun Yu.py compares to Mahmoud's deid_age_mahmoud.py
218        STS Benchmark Score is 0.24266797
219   Shaojun's deid-Shaojun Yu.py compares to ChenLin's deid-ChenLin.py
220        STS Benchmark Score is 0.2644667
221   Shaojun's deid-Shaojun Yu.py compares to ChenLin's deid.py
222        STS Benchmark Score is 0.20311517
223   FeverDreams's deid.py compares to FeverDreams's deid_phone.py
224        STS Benchmark Score is 0.79009724
```

```
225  FeverDreams's deid.py compares to Itoshikihiro's deid-Jie.py
226      STS Benchmark Score is 0.8443497
227  FeverDreams's deid.py compares to Itoshikihiro's deid.py
228      STS Benchmark Score is 0.8443497
229  FeverDreams's deid.py compares to Itoshikihiro's deid-date-Jie.py
230      STS Benchmark Score is 0.63599116
231  FeverDreams's deid.py compares to Itoshikihiro's deid-name-Jie.py
232      STS Benchmark Score is 0.5971642
233  FeverDreams's deid.py compares to Nasim's de-id-location.py
234      STS Benchmark Score is 0.3463645
235  FeverDreams's deid.py compares to Nasim's deid.py
236      STS Benchmark Score is 0.8443497
237  FeverDreams's deid.py compares to Nasim's de-id-date.py
238      STS Benchmark Score is 0.7164128
239  FeverDreams's deid.py compares to Nasim's de-id-age.py
240      STS Benchmark Score is 0.698786
241  FeverDreams's deid.py compares to Andre's deid-Andre_Holder.py
242      STS Benchmark Score is 0.7446203
243  FeverDreams's deid.py compares to Mahmoud's deid_ssn_mahmoud.py
244      STS Benchmark Score is 0.72712636
245  FeverDreams's deid.py compares to Mahmoud's deid_date_mahmoud.py
246      STS Benchmark Score is 0.760545
247  FeverDreams's deid.py compares to Mahmoud's deid_age_mahmoud.py
248      STS Benchmark Score is 0.7525693
249  FeverDreams's deid.py compares to ChenLin's deid-ChenLin.py
250      STS Benchmark Score is 0.73570013
251  FeverDreams's deid.py compares to ChenLin's deid.py
252      STS Benchmark Score is 0.8443497
253  FeverDreams's deid_phone.py compares to Itoshikihiro's deid-Jie.py
254      STS Benchmark Score is 0.75724745
255  FeverDreams's deid_phone.py compares to Itoshikihiro's deid.py
256      STS Benchmark Score is 0.75724745
257  FeverDreams's deid_phone.py compares to Itoshikihiro's deid-date-Jie.py
258      STS Benchmark Score is 0.57370806
259  FeverDreams's deid_phone.py compares to Itoshikihiro's deid-name-Jie.py
260      STS Benchmark Score is 0.54308873
261  FeverDreams's deid_phone.py compares to Nasim's de-id-location.py
262      STS Benchmark Score is 0.34280425
263  FeverDreams's deid_phone.py compares to Nasim's deid.py
264      STS Benchmark Score is 0.75724745
265  FeverDreams's deid_phone.py compares to Nasim's de-id-date.py
266      STS Benchmark Score is 0.6634582
267  FeverDreams's deid_phone.py compares to Nasim's de-id-age.py
268      STS Benchmark Score is 0.65049106
269  FeverDreams's deid_phone.py compares to Andre's deid-Andre_Holder.py
270      STS Benchmark Score is 0.6559067
271  FeverDreams's deid_phone.py compares to Mahmoud's deid_ssn_mahmoud.py
272      STS Benchmark Score is 0.65875167
273  FeverDreams's deid_phone.py compares to Mahmoud's deid_date_mahmoud.py
274      STS Benchmark Score is 0.6717627
275  FeverDreams's deid_phone.py compares to Mahmoud's deid_age_mahmoud.py
276      STS Benchmark Score is 0.66804826
277  FeverDreams's deid_phone.py compares to ChenLin's deid-ChenLin.py
```

```
    STS Benchmark Score is 0.6181232
FeverDreams's deid_phone.py compares to ChenLin's deid.py
    STS Benchmark Score is 0.75724745
Itoshikihiro's deid-Jie.py compares to Itoshikihiro's deid.py
    STS Benchmark Score is 0.9995117
Itoshikihiro's deid-Jie.py compares to Itoshikihiro's deid-date-Jie.py
    STS Benchmark Score is 0.6723099
Itoshikihiro's deid-Jie.py compares to Itoshikihiro's deid-name-Jie.py
    STS Benchmark Score is 0.61155987
Itoshikihiro's deid-Jie.py compares to Nasim's de-id-location.py
    STS Benchmark Score is 0.35966337
Itoshikihiro's deid-Jie.py compares to Nasim's deid.py
    STS Benchmark Score is 0.9995117
Itoshikihiro's deid-Jie.py compares to Nasim's de-id-date.py
    STS Benchmark Score is 0.7326237
Itoshikihiro's deid-Jie.py compares to Nasim's de-id-age.py
    STS Benchmark Score is 0.6962836
Itoshikihiro's deid-Jie.py compares to Andre's deid-Andre_Holder.py
    STS Benchmark Score is 0.7823216
Itoshikihiro's deid-Jie.py compares to Mahmoud's deid_ssn_mahmoud.py
    STS Benchmark Score is 0.7982335
Itoshikihiro's deid-Jie.py compares to Mahmoud's deid_date_mahmoud.py
    STS Benchmark Score is 0.8076906
Itoshikihiro's deid-Jie.py compares to Mahmoud's deid_age_mahmoud.py
    STS Benchmark Score is 0.8005162
Itoshikihiro's deid-Jie.py compares to ChenLin's deid-ChenLin.py
    STS Benchmark Score is 0.6855277
Itoshikihiro's deid-Jie.py compares to ChenLin's deid.py
    STS Benchmark Score is 0.9995117
Itoshikihiro's deid.py compares to Itoshikihiro's deid-date-Jie.py
    STS Benchmark Score is 0.6723099
Itoshikihiro's deid.py compares to Itoshikihiro's deid-name-Jie.py
    STS Benchmark Score is 0.61155987
Itoshikihiro's deid.py compares to Nasim's de-id-location.py
    STS Benchmark Score is 0.35966337
Itoshikihiro's deid.py compares to Nasim's deid.py
    STS Benchmark Score is 0.9995117
Itoshikihiro's deid.py compares to Nasim's de-id-date.py
    STS Benchmark Score is 0.7326237
Itoshikihiro's deid.py compares to Nasim's de-id-age.py
    STS Benchmark Score is 0.6962836
Itoshikihiro's deid.py compares to Andre's deid-Andre_Holder.py
    STS Benchmark Score is 0.7823216
Itoshikihiro's deid.py compares to Mahmoud's deid_ssn_mahmoud.py
    STS Benchmark Score is 0.7982335
Itoshikihiro's deid.py compares to Mahmoud's deid_date_mahmoud.py
    STS Benchmark Score is 0.8076906
Itoshikihiro's deid.py compares to Mahmoud's deid_age_mahmoud.py
    STS Benchmark Score is 0.8005162
Itoshikihiro's deid.py compares to ChenLin's deid-ChenLin.py
    STS Benchmark Score is 0.6855277
Itoshikihiro's deid.py compares to ChenLin's deid.py
    STS Benchmark Score is 0.9995117
```

```
331   Itoshikihiro's deid-date-Jie.py compares to Itoshikihiro's deid-name-Jie.py
332       STS Benchmark Score is 0.7738939
333   Itoshikihiro's deid-date-Jie.py compares to Nasim's de-id-location.py
334       STS Benchmark Score is 0.31878704
335   Itoshikihiro's deid-date-Jie.py compares to Nasim's deid.py
336       STS Benchmark Score is 0.6723099
337   Itoshikihiro's deid-date-Jie.py compares to Nasim's de-id-date.py
338       STS Benchmark Score is 0.6232828
339   Itoshikihiro's deid-date-Jie.py compares to Nasim's de-id-age.py
340       STS Benchmark Score is 0.6010965
341   Itoshikihiro's deid-date-Jie.py compares to Andre's deid-Andre_Holder.py
342       STS Benchmark Score is 0.7292421
343   Itoshikihiro's deid-date-Jie.py compares to Mahmoud's deid_ssn_mahmoud.py
344       STS Benchmark Score is 0.62584037
345   Itoshikihiro's deid-date-Jie.py compares to Mahmoud's deid_date_mahmoud.py
346       STS Benchmark Score is 0.6795608
347   Itoshikihiro's deid-date-Jie.py compares to Mahmoud's deid_age_mahmoud.py
348       STS Benchmark Score is 0.6413636
349   Itoshikihiro's deid-date-Jie.py compares to ChenLin's deid-ChenLin.py
350       STS Benchmark Score is 0.6496881
351   Itoshikihiro's deid-date-Jie.py compares to ChenLin's deid.py
352       STS Benchmark Score is 0.6723099
353   Itoshikihiro's deid-name-Jie.py compares to Nasim's de-id-location.py
354       STS Benchmark Score is 0.36989743
355   Itoshikihiro's deid-name-Jie.py compares to Nasim's deid.py
356       STS Benchmark Score is 0.61155987
357   Itoshikihiro's deid-name-Jie.py compares to Nasim's de-id-date.py
358       STS Benchmark Score is 0.6023469
359   Itoshikihiro's deid-name-Jie.py compares to Nasim's de-id-age.py
360       STS Benchmark Score is 0.60394204
361   Itoshikihiro's deid-name-Jie.py compares to Andre's deid-Andre_Holder.py
362       STS Benchmark Score is 0.66524994
363   Itoshikihiro's deid-name-Jie.py compares to Mahmoud's deid_ssn_mahmoud.py
364       STS Benchmark Score is 0.580127
365   Itoshikihiro's deid-name-Jie.py compares to Mahmoud's deid_date_mahmoud.py
366       STS Benchmark Score is 0.62061316
367   Itoshikihiro's deid-name-Jie.py compares to Mahmoud's deid_age_mahmoud.py
368       STS Benchmark Score is 0.59883475
369   Itoshikihiro's deid-name-Jie.py compares to ChenLin's deid-ChenLin.py
370       STS Benchmark Score is 0.64903986
371   Itoshikihiro's deid-name-Jie.py compares to ChenLin's deid.py
372       STS Benchmark Score is 0.61155987
373   Nasim's de-id-location.py compares to Nasim's deid.py
374       STS Benchmark Score is 0.35966337
375   Nasim's de-id-location.py compares to Nasim's de-id-date.py
376       STS Benchmark Score is 0.35303187
377   Nasim's de-id-location.py compares to Nasim's de-id-age.py
378       STS Benchmark Score is 0.3731488
379   Nasim's de-id-location.py compares to Andre's deid-Andre_Holder.py
380       STS Benchmark Score is 0.33831334
381   Nasim's de-id-location.py compares to Mahmoud's deid_ssn_mahmoud.py
382       STS Benchmark Score is 0.3324147
383   Nasim's de-id-location.py compares to Mahmoud's deid_date_mahmoud.py
```

```
384         STS Benchmark Score is 0.33687252
385  Nasim's de-id-location.py compares to Mahmoud's deid_age_mahmoud.py
386         STS Benchmark Score is 0.34885728
387  Nasim's de-id-location.py compares to ChenLin's deid-ChenLin.py
388         STS Benchmark Score is 0.36347014
389  Nasim's de-id-location.py compares to ChenLin's deid.py
390         STS Benchmark Score is 0.35966337
391  Nasim's deid.py compares to Nasim's de-id-date.py
392         STS Benchmark Score is 0.7326237
393  Nasim's deid.py compares to Nasim's de-id-age.py
394         STS Benchmark Score is 0.6962836
395  Nasim's deid.py compares to Andre's deid-Andre_Holder.py
396         STS Benchmark Score is 0.7823216
397  Nasim's deid.py compares to Mahmoud's deid_ssn_mahmoud.py
398         STS Benchmark Score is 0.7982335
399  Nasim's deid.py compares to Mahmoud's deid_date_mahmoud.py
400         STS Benchmark Score is 0.8076906
401  Nasim's deid.py compares to Mahmoud's deid_age_mahmoud.py
402         STS Benchmark Score is 0.8005162
403  Nasim's deid.py compares to ChenLin's deid-ChenLin.py
404         STS Benchmark Score is 0.6855277
405  Nasim's deid.py compares to ChenLin's deid.py
406         STS Benchmark Score is 0.9995117
407  Nasim's de-id-date.py compares to Nasim's de-id-age.py
408         STS Benchmark Score is 0.7174942
409  Nasim's de-id-date.py compares to Andre's deid-Andre_Holder.py
410         STS Benchmark Score is 0.77435935
411  Nasim's de-id-date.py compares to Mahmoud's deid_ssn_mahmoud.py
412         STS Benchmark Score is 0.69501936
413  Nasim's de-id-date.py compares to Mahmoud's deid_date_mahmoud.py
414         STS Benchmark Score is 0.77950364
415  Nasim's de-id-date.py compares to Mahmoud's deid_age_mahmoud.py
416         STS Benchmark Score is 0.71415687
417  Nasim's de-id-date.py compares to ChenLin's deid-ChenLin.py
418         STS Benchmark Score is 0.65872073
419  Nasim's de-id-date.py compares to ChenLin's deid.py
420         STS Benchmark Score is 0.7326237
421  Nasim's de-id-age.py compares to Andre's deid-Andre_Holder.py
422         STS Benchmark Score is 0.6822848
423  Nasim's de-id-age.py compares to Mahmoud's deid_ssn_mahmoud.py
424         STS Benchmark Score is 0.64283305
425  Nasim's de-id-age.py compares to Mahmoud's deid_date_mahmoud.py
426         STS Benchmark Score is 0.68161535
427  Nasim's de-id-age.py compares to Mahmoud's deid_age_mahmoud.py
428         STS Benchmark Score is 0.7355825
429  Nasim's de-id-age.py compares to ChenLin's deid-ChenLin.py
430         STS Benchmark Score is 0.66115713
431  Nasim's de-id-age.py compares to ChenLin's deid.py
432         STS Benchmark Score is 0.6962836
433  Andre's deid-Andre_Holder.py compares to Mahmoud's deid_ssn_mahmoud.py
434         STS Benchmark Score is 0.7549503
435  Andre's deid-Andre_Holder.py compares to Mahmoud's deid_date_mahmoud.py
436         STS Benchmark Score is 0.8541899
```

```
437  Andre's deid-Andre_Holder.py compares to Mahmoud's deid_age_mahmoud.py
438      STS Benchmark Score is 0.7738091
439  Andre's deid-Andre_Holder.py compares to ChenLin's deid-ChenLin.py
440      STS Benchmark Score is 0.6932734
441  Andre's deid-Andre_Holder.py compares to ChenLin's deid.py
442      STS Benchmark Score is 0.7823216
443  Mahmoud's deid_ssn_mahmoud.py compares to Mahmoud's deid_date_mahmoud.py
444      STS Benchmark Score is 0.7887121
445  Mahmoud's deid_ssn_mahmoud.py compares to Mahmoud's deid_age_mahmoud.py
446      STS Benchmark Score is 0.79113275
447  Mahmoud's deid_ssn_mahmoud.py compares to ChenLin's deid-ChenLin.py
448      STS Benchmark Score is 0.62773407
449  Mahmoud's deid_ssn_mahmoud.py compares to ChenLin's deid.py
450      STS Benchmark Score is 0.79823494
451  Mahmoud's deid_date_mahmoud.py compares to Mahmoud's deid_age_mahmoud.py
452      STS Benchmark Score is 0.81740725
453  Mahmoud's deid_date_mahmoud.py compares to ChenLin's deid-ChenLin.py
454      STS Benchmark Score is 0.67148453
455  Mahmoud's deid_date_mahmoud.py compares to ChenLin's deid.py
456      STS Benchmark Score is 0.8076922
457  Mahmoud's deid_age_mahmoud.py compares to ChenLin's deid-ChenLin.py
458      STS Benchmark Score is 0.6738554
459  Mahmoud's deid_age_mahmoud.py compares to ChenLin's deid.py
460      STS Benchmark Score is 0.8005177
461  ChenLin's deid-ChenLin.py compares to ChenLin's deid.py
462      STS Benchmark Score is 0.6855287
```

Conlcusion

 The average similarity percentage, which is tested by all methods, is above 50%. One of reasons is that all students use one programming language, which is Python. If students had used different programming languages, the average similarity percentage would have been below 50%. Therefore, the similarity scores which are showed in the result section do not represent the true similarity score between two Python files. The similarity checker should ignore the language usage and mainly focus on the algorithm and coding flow.