**Task1**

**Smalldata**

dictionary

[('the', 0), ('of', 1)]

TF-idf matrix

[('AU35', array([3.31979123e-05, 1.89192403e-05, 0.00000000e+00, ...,

0.00000000e+00, 0.00000000e+00, 0.00000000e+00])), ('AU85', array([3.25860691e-05, 1.60795887e-05, 0.00000000e+00, ...,

0.00000000e+00, 0.00000000e+00, 0.00000000e+00]))]

court average TF

[('and', 94.78378378378379), ('applicant', 26.949152542372882), ('court', 17.0), ('attack', 1.6)]

wiki average TF

[('and', 50.50029691211402), ('court', 3.308377896613191), ('attack', 2.953883495145631), ('protein', 5.0227272727272725), ('applicant', 2.4210526315789473)]

**BigData**

dictionary

[('the', 0), ('of', 1)]

TF-idf matrix

[('AU35', array([0.11813274, 0.06732296, 0.00000000e+00, ...,

0.00000000e+00, 0.00000000e+00, 0.00000000e+00])), ('AU85', array([3.25860691e-05, 1.60795887e-05, 0.00000000e+00, ...,

0.00000000e+00, 0.00000000e+00, 0.00000000e+00]))]

court average TF

[('and', 85.8739874319487983), ('applicant', 21.678439935), ('court', 11.5), ('attack', 1.5)]

wiki average TF

[('and', 44.6483902664838), ('court', 2.90889734434), ('attack', 2.5623451321242), ('protein', 7.333333333333), ('applicant', 1.4723947482)]

**Task\_2**

SmallData

1. cost: 24.266477265192247
2. cost: 11.635737959876144
3. cost: 11.406938312529181
4. cost: 11.308633988804733
5. cost: 10.527071971923881
6. cost: 8.346787009720652
7. cost: 8.20417558156651
8. cost: 8.143692490625138
9. cost: 8.195240831113008
10. cost: 7.773769382565333
11. cost: 7.884026215564714
12. cost: 7.90751138486607
13. cost: 5.109729556820676
14. cost: 4.813610704874036
15. cost: 6.389845001075089
16. cost: 3.380113142403782
17. cost: 2.034397652108322
18. cost: 2.699890305435204
19. cost: 1.907226569290628
20. cost: 1.754998203954538

or coefficient: 0.011755358173130281

that coefficient: 0.013820804745077464

not coefficient: 0.0141044940020963

act coefficient: 0.015862577927688203

applicant coefficient: 0.016798012280779873

BigData

0 cost: 24.315580006100316

1 cost: 0.053821008547401177

2 cost: 0.05224861128553595

3 cost: 29.334554592653856

4 cost: 25.416632784601454

5 cost: 11.186842619176755

6 cost: 20.591581230732423

7 cost: 17.126491621633342

8 cost: 21.543625279787637

9 cost: 16.449957979538574

10 cost: 12.190357655459216

11 cost: 10.307995883028784

12 cost: 7.780439805865944

13 cost: 0.49710719840181533

14 cost: 9.960721655963852

15 cost: 8.395065007280667

16 cost: 7.86378873347186

17 cost: 2.165556479031405

18 cost: 4.905996631578893

19 cost: 0.8336352830351362

20 cost: 9.403788106660645

21 cost: 11.585802692896571

22 cost: 7.4362223090940835

23 cost: 1.930675013516446

24 cost: 0.5963941780558321

25 cost: 14.62299468215285

26 cost: 5.257948536811272

27 cost: 8.56668534606305

28 cost: 2.26408381117439

29 cost: 13.571875765070969

30 cost: 7.917601217928234

31 cost: 8.129427771468203

32 cost: 8.056891519823898

33 cost: 4.716583821249995

34 cost: 11.720552477437478

35 cost: 7.167296753969953

36 cost: 9.422995276173205

37 cost: 1.5925123537743833

38 cost: 9.099982991022122

39 cost: 11.157319154794594

40 cost: 8.276970471684924

41 cost: 11.405645612221212

42 cost: 9.786401417532764

43 cost: 0.7344871599940395

44 cost: 13.9806747972432

45 cost: 5.95604907486958

46 cost: 8.50830436100278

47 cost: 7.569928343271928

48 cost: 2.6823665821101055

49 cost: 4.892012222442395

**Task3**

SmallData

TP: 74

FP: 2332

FN: 0

TN: 1036

F1: 0.05967741935483871

BigData

TP: 74

FP: 3216

FN: 0

TN: 152

F1: 0.04399524375743163

Screenshot





