Afsah ur Rehman
Bscs22119

Assignment 1

Problems

I tried using multi processing metal got some issues while syncing them and my hardware was good enough so I don't know how to remove the multi threading from the code To get rid of the async issues,

Dataset:

I used a smaller version of data set which was first seventeen txt that and I reduced it only to seven txts from the original named as small_dataset
Also I haven't uploaded the complete dataset as it was a large file.

Results

Also available in json file


```
{
  "LDA": [
    [
      "W13-1914.pdf.txt",
      0.07279339846152835
    ],
    [
      "N10-1070.pdf.txt",
      0.0621967728061531
    ],
    [
      "D09-1026.pdf.txt",
      0.060749930837848534
    ],
    [
      "W11-1102.pdf.txt",
      0.0549212212574603 65
    ],
    [
      "P07-1085.pdf.txt",
```

```
      0.05322274085793983
    ]
  ],
  "Topic modelling": [
    [
      "W06-1201.pdf.txt",
      0.0759641462762091
    ],
    [
      "W10-4104.pdf.txt",
      0.06835377557034689
    ],
    [
      "W14-3112.pdf.txt",
      0.06670296625903069
    ],
    [
      "W11-0801.pdf.txt",
      0.06603650365615214
    ],
    [
      "P12-1079.pdf.txt",
      0.06315509738195756
    ]
  ],
  "Generative models": [
    [
      "P86-1028.pdf.txt",
      0.03314458436039147
    ],
    [
      "W09-3839.pdf.txt",
      0.03268775434169865
    ],
    [
      "W06-1668.pdf.txt",
      0.03185093731639403
    ],
    [
```

```
      "W04-0303.pdf.txt",
      0.03074289946840773
    ],
    [
      "D15-1231.pdf.txt",
      0.029145637087746593
    ]
  ],
  "Semantic relationships between terms": [
    [
      "W15-2616.pdf.txt",
      0.0461396276965222
    ],
    [
      "W04-1801.pdf.txt",
      0.04396616625869506
    ],
    [
      "W04-1117.pdf.txt",
      0.03825964858721889
    ],
    [
      "W09-2004.pdf.txt",
      0.037611696434265575
    ],
    [
      "P10-5006.pdf.txt",
      0.0349203500282124
    ]
  ],
  "Natural Language Processing": [
    [
      "W15-4800.pdf.txt",
      0.024053988110256067
    ],
    [
      "D15-1192.pdf.txt",
      0.019243190488204855
    ],
```

```
    [
      "W12-0506.pdf.txt",
      0.01835056693484171
    ],
    [
      "W06-1645.pdf.txt",
      0.01781776897056005
    ],
    [
      "A83-1033.pdf.txt",
      0.01602994286954911
    ]
  ],
  "Text Mining": [
    [
      "W04-3107.pdf.txt",
      0.09254156939449391
    ],
    [
      "J08-1004.pdf.txt",
      0.0794753044025628
    ],
    [
      "W14-1101.pdf.txt",
      0.0571801566990369
    ],
    [
      "J11-1012.pdf.txt",
      0.05603544416342162
    ],
    [
      "W04-3109.pdf.txt",
      0.04720241123477542
    ]
  ],
  "Translation model": [
    [
      "A97-2006.pdf.txt",
      0.06199220796893988
```

```
    ],
    [
      "W15-4111.pdf.txt",
      0.055753834198591826
    ],
    [
      "W06-0801.pdf.txt",
      0.051728154783551666
    ],
    [
      "W02-1600.pdf.txt",
      0.048503793487785926
    ],
    [
      "W15-4110.pdf.txt",
      0.04583301613641186
    ]
  ],
  "Learning procedures for the lexicon": [
    [
      "W10-2505.pdf.txt",
      0.04708940245489384
    ],
    [
      "W99-0500.pdf.txt",
      0.04403164904873189
    ],
    [
      "W00-1801.pdf.txt",
      0.04042249748736043
    ],
    [
      "W10-3401.pdf.txt",
      0.037362927626480164
    ],
    [
      "P06-2099.pdf.txt",
      0.03558314117300646
    ]
```

```
  ],
  "Semantic evaluations": [
    [
      "P10-5006.pdf.txt",
      0.0349203500282124
    ],
    [
      "W08-2201.pdf.txt",
      0.03377712428324116
    ],
    [
      "P81-1008.pdf.txt",
      0.03291286082431361
    ],
    [
      "W97-0218.pdf.txt",
      0.027919099038546197
    ],
    [
      "W06-1419.pdf.txt",
      0.027094337214301028
    ]
  ],
  "System results and combination": [
    [
      "W11-2121.pdf.txt",
      0.03145651564460771
    ],
    [
      "W12-5702.pdf.txt",
      0.028927189429781338
    ],
    [
      "W14-1001.pdf.txt",
      0.026475774542753107
    ],
    [
      "N10-1141.pdf.txt",
      0.024218496815249076
```

```
    ],
    [
      "P11-1127.pdf.txt",
      0.02258540318842611
    ]
  ]
}
```