

DUE: 2/26/2017 11:59pm EST Total: 10 points

1) Term-Document Matrix

This turned out to be a large matrix when outputted into .csv and it is largely useless because of its sparsity. See screenshots below for a sample. Matrix will be attached, but it is 86.2 MB.

The code for making the conversion will be attached (made using python). Limited screenshots available below as well.

The screenshot shows an Excel spreadsheet titled 'term_document_matrix_MP.xlsx'. The spreadsheet contains a large matrix of 0s and 1s. The columns are labeled with words: 'orthogonal', 'yellow', 'interchan', 'prefix', 'lori', 'lord', 'digit', 'uofa', 'hough', 'saskatch', 'dell', 'politician', 'fractal', 'viewabl', 'disturb', 'prize', 'wooden', 'jessup', 'demmel', 'terabyt', 'wednesdi', 'charter', 'navath', 'miller', 'frederick', 'histor', 'second', 'valiant', 'ku'. The rows are labeled 'Train' and numbered 1 through 38. A green column on the left is labeled 'matrix' and a green row at the top is labeled 'index'. The spreadsheet interface includes the standard Excel ribbon with tabs like File, Home, Insert, Page Layout, Formulas, Data, Review, View, Add-ins, Font Reader PDF, and Team. The status bar at the bottom indicates 'Ready'.

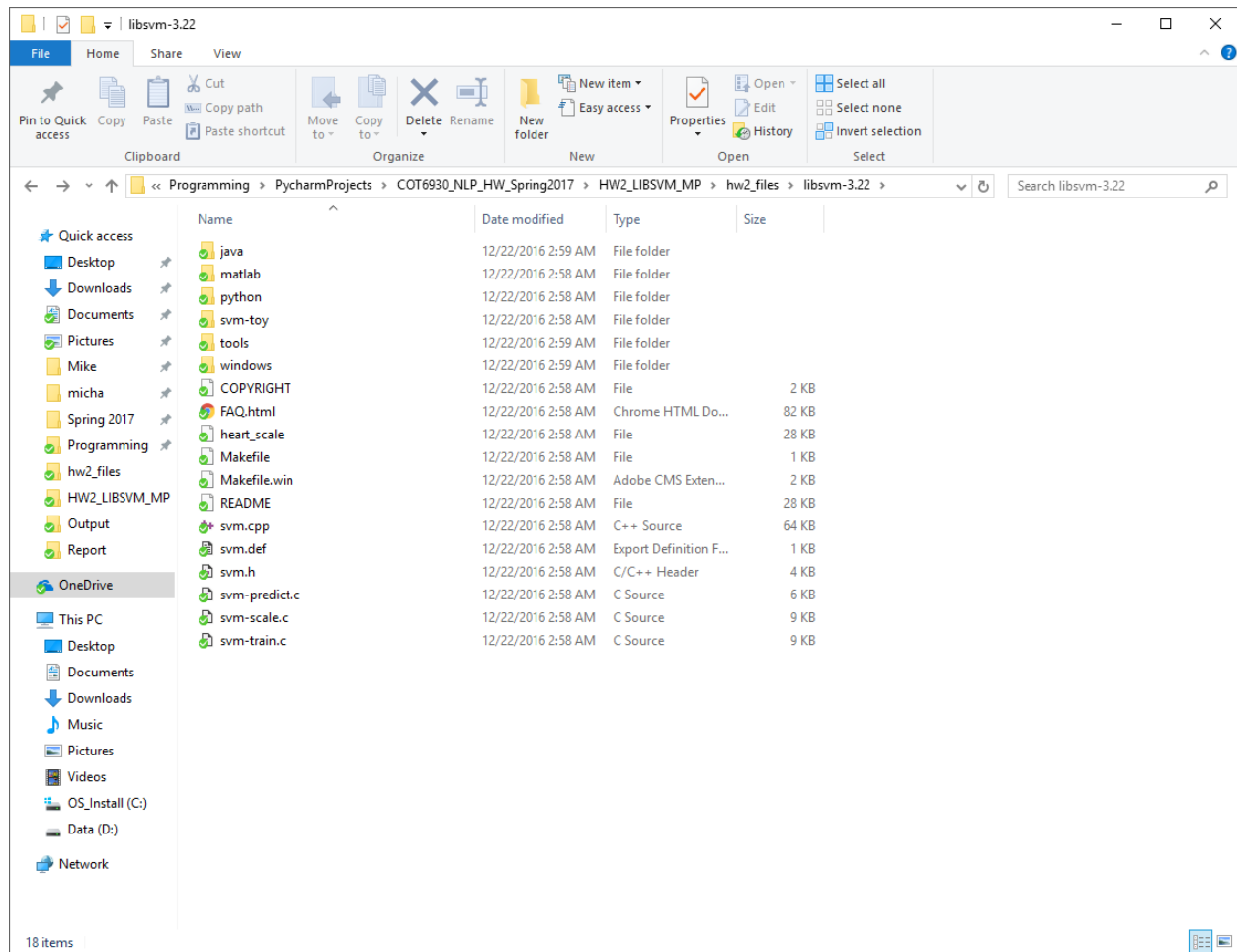
[illegible]

Michael Perez
Homework 2

The image displays a PyCharm IDE environment with two main windows. The top window is an Excel spreadsheet titled 'term_document_matrix_MP.xlsx'. It contains a large table with rows labeled '4164 Test' through '4201 Test' and columns labeled 'A' through 'AC'. The cells contain numerical values, mostly 0, with some 1s and 2s, representing a binary or weighted matrix. The bottom window is a Python script named 'test_to_libsvm.py' located at 'C:\Users\micha\OneDrive\Documents\Programming\PycharmProjects\COT6930_NLP_Spring2017\HW2_LIBSVM_MP\test_to_libsvm.py'. The script is a Jupyter Notebook-style file with code cells. It defines functions for reading text files, processing them into a term-document matrix, and saving the results. The code includes comments in Spanish and English, and uses libraries like 'textmining' and 'libsvm'. The script is currently running, as indicated by the 'Run' button and the output console at the bottom.

```
1 # COT6930_NLP_HW_Spring2017
2 # HW2_LIBSVM_MP
3 # test_to_libsvm.py
4
5 # Importing libraries
6 import textmining
7
8 # Defining functions
9 def class_label_to_number(class_label):
10     return {'Bad': 0, 'Good': 1, 'Neutral': 2, 'Calculation': 3, 'Exploratory': 4, 'Input': 5, 'Linked Cell': 6, 'Note': 7}
11
12 def read_file_text(file_path, libsvm, docs, tdm):
13     # Create lists for input and output file paths.
14     input_paths = []
15     output_paths = []
16
17     # Insert file paths into lists.
18     input_paths.append('../hw2_files/Datasets/webkb-train-stemmed.txt')
19     input_paths.append('../hw2_files/Datasets/webkb-test-stemmed.txt')
20     output_paths.append('../hw2_files/Output/webkb-train-libsvm.txt')
21     output_paths.append('../hw2_files/Output/webkb-test-libsvm.txt')
22
23     # Lists containing documents separated by training and testing file
24     docs_tr = []
25     docs_te = []
26
27     # Lists that will contain the libsvm formatted documents separated by training and testing file
28     libsvm_tr = []
29     libsvm_te = []
30
31     # Initialize class to create term-document matrix
32     tdm = textmining.TermDocumentMatrix()
33
34     # Use function to process files
35     read_file_text(input_paths[0], libsvm_tr, docs_tr, tdm)
36     read_file_text(input_paths[1], libsvm_te, docs_te, tdm)
37
38     # Offset for test data output
39     offset_te = len(docs_tr)
40     # get tdm rows -> MAKE SURE TO EDIT textmining __init__.py as instructed above
41     tdm_rows = tdm.rows(offset=offset_te)
42
43     # Converting training and testing text data to libsvm format
44     for tdm_row_index, row in enumerate(tdm_rows):
45         # Output files to specified path
46         with open(output_paths[0], 'w') as output_file:
47             with open(output_paths[1], 'w') as output_file:
48                 # ... (code continues) ...
```

2) LIBSVM Download



3) LIBSVM Formatted Text

Attached files under the following names “webkb-test-libsvm.txt” and “webkb-train-libsvm.txt”.

Screenshots can be found below as well:

webkb-test-libsvm.txt

```
webkb-test-libsvm.txt - Notepad
File Edit Format View Help
1 750:1 1822:1 1893:1 2513:1 3055:1 3826:2 5464:1
4 186:1 242:1 286:1 843:1 966:3 1188:4 1299:1 1452:1 1493:1 1680:2 1690:1 2034:1 2146:2 2328:1 2446:3 2611:1 2819:1 3103:1 3228:1 3279:1 3364:1 3640:1 3716:3 3727:1 3762:1 37
1 145:1 198:2 212:1 430:1 448:1 599:2 723:1 750:3 798:1 919:1 1145:1 1161:1 1188:1 1434:3 1439:1 1690:2 1766:1 1870:1 1890:2 1936:1 2079:4 2164:1 2354:1 2484:1 2697:2 2700:1
1 101:1 318:1 359:1 430:1 568:2 668:3 773:1 867:1 946:1 1188:4 1289:1 1353:1 1395:1 1434:2 1493:1 1536:1 1545:1 1643:1 1680:1 1734:1 1804:3 1847:2 1850:1 1890:3 2290:5 2374:3
4 364:1 569:1 628:1 661:1 676:1 1081:1 1188:1 1360:1 1395:2 1469:1 1831:1 2135:1 2164:1 2627:2 2883:2 3354:1 3526:1 3870:1 4573:1 4853:1 4888:1 5673:1 5935:2 6407:1 6430:1 64
2 134:1 268:2 584:1 695:1 771:3 898:1 1036:1 1042:5 1081:1 1145:1 1188:10 1323:2 1469:6 1690:1 1715:2 1776:1 1801:1 2164:1 2174:2 2512:1 2546:1 2596:2 2620:1 2773:1 327
4 60:1 120:2 212:5 231:4 242:1 331:1 364:1 407:5 475:1 536:7 628:5 681:1 692:1 771:1 774:4 1017:1 1161:1 1188:2 1198:1 1287:1 1365:1 1374:2 1432:1 1479:1 1545:1 1553:1 1608:1
1 33:3 52:1 65:4 71:1 72:1 73:1 87:2 116:1 134:1 144:1 152:2 154:1 160:1 196:1 212:1 242:1 293:1 331:1 452:1 504:2 535:1 549:2 594:1 609:1 643:1 668:3 698:1 723:3 753:1 772:1
674:1 4720:1 4737:1 4782:2 4836:1 4837:1 4839:1 4843:1 4849:1 4890:1 4904:1 4926:2 4972:2 4977:1 4978:3 5009:1 5016:1 5080:1 5118:1 5129:1 5201:1 5203:1 5298:1 5412:1 5426:1
2 242:1 247:1 504:1 558:1 1017:1 1042:2 1067:5 1145:4 1188:2 1345:1 1395:3 1492:1 1690:3 1761:1 1785:1 1847:1 1959:1 1990:1 2071:1 2074:1 2096:2 2214:1 2374:1 2409:1 2444:3 2
3 72:1 152:1 212:1 231:1 536:1 558:1 582:1 695:1 1042:1 1145:3 1188:2 1581:2 1680:4 1785:1 2079:1 2115:1 2192:1 2522:2 2596:1 2709:3 2819:3 3088:1 3421:1 3716:2 4255:3 4304:1
1 45:1 198:1 312:1 430:2 599:4 668:1 755:1 1058:1 1145:1 1188:4 1433:1 1509:2 1545:1 1680:2 1690:1 1695:2 1827:1 2079:2 2113:1 2347:1 2490:1 2547:3 2562:1 2565:1 2876:2 3004:
2 37:1 127:1 152:1 212:1 231:1 242:2 348:1 364:1 407:3 430:2 444:1 448:2 503:7 512:2 524:1 549:2 558:1 592:1 654:1 659:1 712:2 725:1 761:2 771:1 772:1 789:1 898:1 903:1 942:1
024:1 6150:1 6158:1 6160:1 6384:1 6427:1 6430:1 6452:1 6473:1 6555:1 6699:15 6777:1 6826:1 6860:1 6837:2 6839:1 6840:1 6865:1 6873:11 6945:1 6957:1 6968:1 7000:1 7028:1 7087:
1 549:1 558:2 919:1 1785:1 3147:1 3919:1 4130:1 4366:1 4420:1 4983:1 5071:1 6706:1 6823:1 6835:1 7148:1 7338:1
4 40:4 166:1 861:1 1188:1 1581:1 1690:1 1785:1 1959:1 1988:1 2146:1 2347:1 2473:2 2500:2 2550:1 2709:1 3304:1 3310:6 3421:2 3486:1 3638:1 3716:1 3842:1 3919:1 4430:1 4839:2 5
2 7:1 152:2 212:3 242:2 300:1 365:1 407:1 504:5 517:1 520:1 535:1 617:1 628:1 654:1 746:1 771:1 850:1 867:1 928:1 967:2 1030:1 1032:3 1033:1 1036:3 1042:1 1109:1 1141:1 1145:
884:1 6908:1 6934:2 6953:1 6965:1 6979:1 7001:1 7028:1 7126:2 7135:1 7166:1 7189:1 7305:1 7347:1 7553:1 7580:1 7628:2 7660:1 7667:1 7694:1 7715:1
4 21:5 56:2 242:1 738:2 774:1 843:1 884:1 1013:1 1036:1 1120:1 1188:2 1217:4 1251:2 1253:1 1434:1 1865:2 1959:6 1964:1 1984:1 2021:1 2189:1 2251:2 2285:1 2310:2 2484:3 2512:1
4 212:1 558:1 836:1 837:26 1188:1 1217:10 1311:1 1365:3 1988:2 2146:4 2446:15 2613:2 2803:41 3071:1 3103:1 3749:15 4035:1 4230:1 4277:1 4524:1 4749:1 4839:3 5625:1 5834:3 2512:1
5 12:1 107:1 145:1 247:6 257:1 263:1 430:3 448:1 525:1 549:2 654:3 755:1 771:1 772:4 774:1 851:2 852:3 869:1 874:1 891:2 898:1 1058:1 1067:1 1145:2 1188:3 1198:1 1241:1 1307:
1 128:1 1042:1 1067:1 1145:1 1345:1 1395:1 1888:1 2065:1 2079:1 2146:1 2622:1 2992:1 3842:1 4164:2 4322:2 4765:1 5045:1 5313:1 5462:1 5601:1 5696:1 5776:1 6339:1 6817:1 6997:
3 59:1 72:2 123:1 137:1 152:1 208:1 212:6 229:4 299:3 300:1 311:1 397:3 422:1 430:1 478:2 504:1 525:1 531:5 536:1 592:2 628:3 661:1 669:7 703:1 774:3 842:1 872:2 889:1 908:4
196:1 7525:1 7335:2 7570:1 7597:3 7632:1 7762:2
1 107:1 145:2 152:2 198:2 208:4 212:1 263:3 405:3 407:1 448:1 486:1 525:1 723:1 772:1 874:1 898:1 991:1 1017:4 1036:1 1085:4 1109:1 1145:8 1188:5 1307:2 1323:1 1374:4 1558:2
4 617:2 1078:3 1217:1 1244:1 1374:1 1493:1 1785:1 1890:2 2021:1 2251:1 2302:2 2583:1 2887:1 3364:3 3749:2 3762:2 3771:1 3789:1 3795:2 3959:2 4006:1 4518:1 4524:3 4575:1 4888:
1 65:1 72:1 79:1 212:1 218:1 288:2 466:1 504:1 524:1 723:4 865:1 1036:1 1188:4 1357:1 1434:2 1479:1 1785:1 2079:1 2115:2 2151:1 2446:1 2525:1 2566:1 2640:1 2683:1 2709:1 2783
4 21:1 56:1 200:1 212:1 504:1 547:1 750:1 1287:1 1801:2 1984:1 2397:1 2591:1 2613:1 3103:1 3115:1 3194:1 3364:1 3562:1 3822:1 4554:2 4888:1 5508:2 5601:1 5895:1 6430:1 6494:1
2 186:2 314:1 397:1 524:1 1013:1 1033:1 1036:1 1078:1 1145:1 1188:6 1307:3 1323:3 1410:1 1643:2 1690:1 1802:1 1937:1 2317:1 2395:1 2547:3 2554:1 2614:1 2709:5 3098:2 3364:1 3
2 7:2 72:2 137:1 212:1 242:1 254:4 346:1 365:1 407:2 430:2 517:3 558:1 772:2 898:3 1036:2 1109:2 1145:2 1167:1 1188:3 1206:1 1298:1 1437:3 1608:1 1680:6 1761:1 1785:1 1890:1
2 7:1 152:2 242:1 771:1 1044:2 1145:1 1149:1 1690:1 2063:1 3762:1 3805:1 4164:1 4256:1 4327:1 4430:3 4524:1 4990:1 5265:1 5836:1 6122:1 6510:1 6591:1 6617:1 6689:1 6873:3 745
4 21:2 27:1 30:1 56:1 72:1 74:2 79:1 83:1 98:2 101:1 128:2 134:2 167:1 186:1 211:2 212:1 216:2 229:1 231:3 242:1 258:1 263:1 268:1 283:3 284:1 285:1 299:1 311:7 313:1 364:3 3
:1 2164:1 2172:4 2184:1 2192:3 2196:2 2229:1 2249:1 2251:1 2275:1 2302:2 2310:1 2364:1 2410:3 2416:1 2444:3 2460:4 2470:3 2484:1 2486:1 2488:2 2495:1 2499:2 2504:1 2512:3 252
53:1 4255:1 4259:1 4284:1 4300:2 4302:1 4304:1 4322:1 4365:2 4396:1 4446:4 4447:2 4483:1 4495:3 4508:1 4518:1 4521:4 4523:2 4524:2 4542:4 4547:2 4558:4 4570:9 4573:8 4574:1 4
:5 6340:6 6380:1 6407:3 6429:1 6430:5 6437:2 6452:2 6453:1 6461:1 6473:1 6483:1 6497:1 6502:2 6509:1 6515:1 6521:1 6523:3 6555:2 6575:1 6585:1 6588:2 6591:7 6629:4 6637:3 666
2 145:1 150:1 152:1 212:1 242:2 288:7 723:4 908:1 1036:3 1145:2 1188:9 1238:1 1287:3 1307:1 1583:1 1680:2 1785:1 1890:1 2115:2 2164:1 2229:1 2328:1 2552:1 2640:1 2658:1 2703:
2 60:1 105:1 212:1 242:4 341:1 387:3 504:1 524:1 594:1 698:1 702:1 852:2 896:1 898:1 1036:1 1042:1 1145:3 1181:1 1188:3 1307:1 1545:2 1690:1 1785:1 1801:3 1837:1 1926:1 2097:
4 7:1 152:1 163:1 166:1 364:1 504:1 577:1 628:3 654:1 874:1 1036:2 1188:2 1287:1 1345:1 1357:1 1387:1 1531:1 1606:3 1643:3 1690:1 1889:1 1890:2 1988:1 2115:5 2118:1 2146:1 21

```

webkb-train-libsvm.txt

```
webkb-train-libsvm.txt - Notepad
File Edit Format View Help
1 21:1 60:1 101:1 107:1 166:1 208:3 243:1 331:2 453:1 577:1 654:1 661:1 677:7 688:1 692:1 732:1 761:2 852:1 867:1 995:1 1036:1 1076:1 1109:1 1123:1 1145:1 1188:11 1283:2 1287
7502:1 7587:1 7607:1 7625:1 7667:1 7762:1 7767:1
1 73:1 628:1 707:1 865:1 919:1 1133:1 1188:1 1890:3 2115:2 2146:2 2183:1 2613:1 2750:1 2870:1 3228:1 3762:1 4065:1 4524:1 5166:1 5314:1 5731:1 6252:2 6339:1
2 21:1 211:1 242:1 524:1 669:1 723:1 908:1 1004:1 1036:1 1047:1 1050:1 1096:1 1108:1 1189:1 1145:2 1169:1 1188:3 1353:1 1469:3 1558:1 1642:1 1785:1 1875:1 1959:3 1984:1 2032:
1 231:1 338:1 407:1 525:4 558:2 867:1 868:1 886:1 1032:3 1062:2 1042:3 1067:1 1145:3 1163:1 1188:4 1244:1 1307:1 1434:5 1469:1 1602:1 1623:1 1690:3 1758:1 1785:1 1870:1 1959:
3 212:1 331:1 443:1 549:1 774:1 867:1 1036:1 1067:3 1145:3 1188:2 1345:1 1410:3 1545:1 1581:1 1690:2 1890:2 1997:1 2079:1 2114:1 2152:2 2395:2 2566:1 2614:1 3071:1 3263:1 329
2 242:1 430:1 549:1 765:1 1036:1 1188:1 1188:1 1848:1 2146:1 2505:1 3113:1 3151:2 3716:1 4430:1 4440:1 4888:1 4890:1 5107:1 5895:1 5952:1 6510:1 6840:1
2 242:1 524:2 558:1 852:1 1017:1 1036:1 1042:1 1109:3 1188:2 1375:1 1770:2 1785:1 2353:1 2446:1 2568:1 2862:1 3042:1 3342:1 3364:2 3397:1 3716:3 3863:1 3942:1 4430:2 4477:2 4
2 107:1 242:2 341:1 517:1 558:1 577:1 599:4 628:1 723:1 784:4 954:2 1017:1 1036:1 1145:6 1188:7 1203:1 1345:1 1365:1 1418:1 1487:1 1514:1 1524:2 1544:1 1553:1 1608:2 1690:1 1
1 8:1 73:1 109:1 212:1 867:1 1036:1 1109:1 1163:1 1188:2 1199:1 1434:3 1680:1 1747:1 2056:1 2146:2 2354:1 2512:2 2617:1 2622:1 2639:1 2709:1 2803:1 3217:1 3279:1 3301:1 3716:
4 21:2 212:1 426:1 524:1 531:1 601:1 667:1 918:1 1098:1 1244:1 1287:1 1469:1 1531:1 1964:1 1984:1 2131:2 2146:4 2161:1 2512:2 2587:1 2614:1 2764:1 3076:1 3103:4 3275:4 3364:1
1 102:1 212:1 595:2 628:3 1036:1 1188:1 1514:1 1533:1 1581:1 1952:2 2076:1 2110:1 2270:2 2613:1 2697:3 2700:2 3420:1 3643:1 3716:2 3724:1 3860:1 3942:1 4032:1 4188:2 4230:1 4
2 101:1 212:1 242:1 263:3 549:1 558:2 573:1 868:1 1036:1 1050:1 1106:1 1145:4 1163:4 1188:4 1244:1 1287:1 1342:1 1675:1 1690:1 1785:2 2146:4 2416:1 2512:3 2583:1 2839:1 2888:
1 300:1 524:2 599:1 661:1 723:3 750:1 1011:3 1036:1 1145:1 1163:1 1188:4 1307:1 1353:1 1394:1 1451:1 1479:1 1558:1 1583:1 1690:2 1785:1 1870:1 2058:3 2079:1 2251:1 2328:1 262
1 80:1 93:1 293:2 331:1 504:2 654:1 658:1 750:9 774:1 786:1 1013:1 1067:2 1131:3 1188:4 1377:1 1434:4 1479:2 1493:1 1582:1 1583:1 1693:1 1704:2 1745:2 1783:1 1847:2 1850:1 18
2 211:1 231:6 242:1 407:1 504:1 628:1 692:2 774:2 1145:1 1188:1 1217:1 1256:1 1374:2 1469:1 1623:2 1680:2 1690:1 2139:1 2444:1 2451:1 2709:3 2731:4 2754:1 2868:1 3087:1 3229:
2 72:1 212:2 242:3 427:1 430:2 475:1 504:1 549:1 621:1 654:2 664:1 761:1 857:1 874:1 903:1 1036:2 1044:1 1058:1 1141:1 1145:4 1188:6 1307:1 1345:1 1437:1 1545:3 1581:3 1627:1
3 35:1 212:1 231:1 242:1 266:1 288:2 430:1 531:1 661:6 679:1 692:1 851:1 1036:2 1041:1 1127:1 1145:3 1188:3 1240:1 1357:1 1387:1 1434:1 1469:1 1581:1 1583:3 1651:1 1690:3 178
1 59:1 198:2 212:1 309:3 504:1 558:1 626:1 867:1 1188:3 1244:1 1622:1 1646:1 1747:1 1785:2 1990:1 2146:1 2186:1 2217:1 2388:1 2399:2 2640:1 2676:1 2686:2 2750:2 2876:1 3074:1
2 107:1 211:1 231:1 242:1 311:3 425:1 504:3 723:1 730:1 738:1 908:1 967:1 1017:1 1036:1 1042:1 1047:2 1086:2 1145:1 1188:6 1295:2 1323:2 1345:1 1374:1 1575:7 1690:2 1856:1 20
1 111:3 152:1 198:1 218:1 430:3 518:1 599:2 991:1 1141:1 1145:3 1188:4 1198:1 1289:3 1434:2 1473:1 1521:1 1545:1 1581:1 1747:1 1804:3 2000:1 2054:1 2079:2 2130:2 2194:1 2290:
4 242:1 311:1 475:1 504:12 524:1 600:1 772:1 987:1 1013:1 1042:2 1058:1 1244:1 1351:1 1389:1 1623:1 1690:1 1715:6 1959:1 2079:1 2146:2 2300:2 2451:4 2566:1 2614:2 2638:1 30
3 72:1 152:1 175:3 594:1 1145:2 1847:1 2164:2 2418:1 2764:1 2796:1 2990:1 4430:1 4524:1 4707:1 4782:1 4917:1 5553:1 5684:1 6043:1 6160:1
1 198:1 212:1 240:1 312:1 558:1 1036:1 1188:1 1524:1 1690:1 2111:1 2355:1 2709:1 2839:1 3716:1 3762:1 3771:1 3889:1 4025:1 4130:1 4291:1 4352:1 4366:2 4430:1 4431:1 5085:1 58
2 107:7 198:1 229:1 242:2 259:1 311:2 524:3 724:1 791:1 827:1 898:2 1004:2 1020:3 1036:1 1096:1 1109:3 1145:1 1188:9 1295:1 1323:1 1345:1 1469:1 1608:1 1643:1 1715:1 1749:1 1
2 152:2 212:1 242:1 341:3 365:1 430:1 486:1 517:1 599:1 607:1 654:1 738:1 772:1 773:1 908:1 1036:3 1090:1 1145:2 1188:8 1246:1 1323:3 1418:1 1435:1 1469:3 1520:1 1542:1 1545:
1 617:1 1785:1 2602:1 3103:3 3299:1 3456:2 3964:1 4706:3 6639:1 7325:2
2 70:3 87:1 103:1 212:2 231:1 242:3 311:2 350:1 407:1 549:1 617:1 628:1 654:1 661:1 738:1 771:2 782:1 784:8 792:1 867:1 903:2 908:1 954:1 967:1 997:1 1008:1 1036:1 1042:2 107
870:1 5935:3 5973:3 6051:12 6081:1 6092:4 6234:2 6257:1 6312:2 6330:12 6399:1 6411:1 6507:1 6510:6 6555:1 6591:1 6618:1 6689:3 6770:1 6776:1 6777:1 6837:1 6841:1 6873:1 6908:
2 17:1 629:1 1188:4 1404:1 1434:1 1544:1 1608:1 1680:1 1890:1 1959:1 2085:1 2054:2 2172:1 2229:1 2394:2 2405:1 2598:4 2839:1 3562:1 3716:1 3872:1 3920:1 4523:1
2 10:1 212:1 242:2 462:1 549:1 668:1 695:1 1044:1 1076:1 1109:1 1145:2 1188:1 1353:1 1357:1 1714:1 1890:1 2115:1 2419:1 2566:1 2649:3 2724:1 2741:1 2796:1 3076:1 3342:1 3578:
2 17:1 242:1 355:1 508:4 517:1 525:2 549:1 554:1 628:2 693:2 697:3 718:1 780:1 827:1 850:1 899:1 1017:5 1036:1 1042:1 1067:1 1145:2 1155:2 1188:6 1235:1 1239:1 1291:5 1312:4
1 46:1 203:1 285:1 405:1 594:1 761:3 818:1 1036:1 1145:2 1188:8 1307:1 1583:1 1690:1 1785:1 1868:1 2079:1 2259:2 2328:1 2512:1 2709:1 2839:1 3098:1 3716:3 3822:1 3848:1 3959:
1 45:3 72:2 101:1 594:1 668:1 705:1 750:1 812:2 1036:2 1163:1 1188:2 1579:3 1847:3 2079:2 1800:1 2221:1 2383:1 2392:1 2409:1 2451:5 2484:1 2505:3 2537:1 2587:4 2683:4 2803:4
4 105:1 152:1 170:1 212:1 324:1 524:1 569:1 774:1 784:3 942:1 1008:1
```

4) LIBSVM Training and Testing Data Demo

svm-train.exe -t 0 webkb-train-libsvm.txt

```
Command Prompt
Microsoft Windows [Version 10.0.14393]
(c) 2016 Microsoft Corporation. All rights reserved.

C:\Users\micha>cd C:\Users\micha\OneDrive\Documents\Programming\PycharmProjects\COT6930_NLP_HW_Spring2017\HW2_LIBSVM_MP\hw2_files\libsvm-3.22\windows

C:\Users\micha\OneDrive\Documents\Programming\PycharmProjects\COT6930_NLP_HW_Spring2017\HW2_LIBSVM_MP\hw2_files\libsvm-3.22\windows>svm-train.exe -t 0 webkb-train-libsvm.txt
...*
optimization finished, #iter = 3566
nu = 0.017010
obj = -23.405295, rho = -0.999908
nSV = 502, nBSV = 17
.*
optimization finished, #iter = 1442
nu = 0.009897
obj = -9.091582, rho = -0.999766
nSV = 279, nBSV = 5
.*
optimization finished, #iter = 1620
nu = 0.009250
obj = -9.278550, rho = -0.999735
nSV = 257, nBSV = 5
.*
optimization finished, #iter = 1817
nu = 0.015540
obj = -11.544269, rho = -0.999805
nSV = 287, nBSV = 10
.*
optimization finished, #iter = 1239
nu = 0.012987
obj = -11.896631, rho = -0.999740
nSV = 202, nBSV = 10
.*
optimization finished, #iter = 934
nu = 0.010839
obj = -7.572709, rho = -0.216789
nSV = 178, nBSV = 5
Total nSV = 1109

C:\Users\micha\OneDrive\Documents\Programming\PycharmProjects\COT6930_NLP_HW_Spring2017\HW2_LIBSVM_MP\hw2_files\libsvm-3.22\windows>
```

svm-predict.exe webkb-test-libsvm.txt webkb-train-libsvm.txt.model webkb_output_file_MP

```
C:\Users\micha\OneDrive\Documents\Programming\PycharmProjects\COT6930_NLP_HW_Spring2017\HW2_LIBSVM_MP\hw2_files\libsvm-3.22\windows>svm-predict.exe webkb-test-libsvm.txt webkb
-train-libsvm.txt.model webkb_output_file_MP
Accuracy = 85.0287% (1187/1396) (classification)

C:\Users\micha\OneDrive\Documents\Programming\PycharmProjects\COT6930_NLP_HW_Spring2017\HW2_LIBSVM_MP\hw2_files\libsvm-3.22\windows>
```