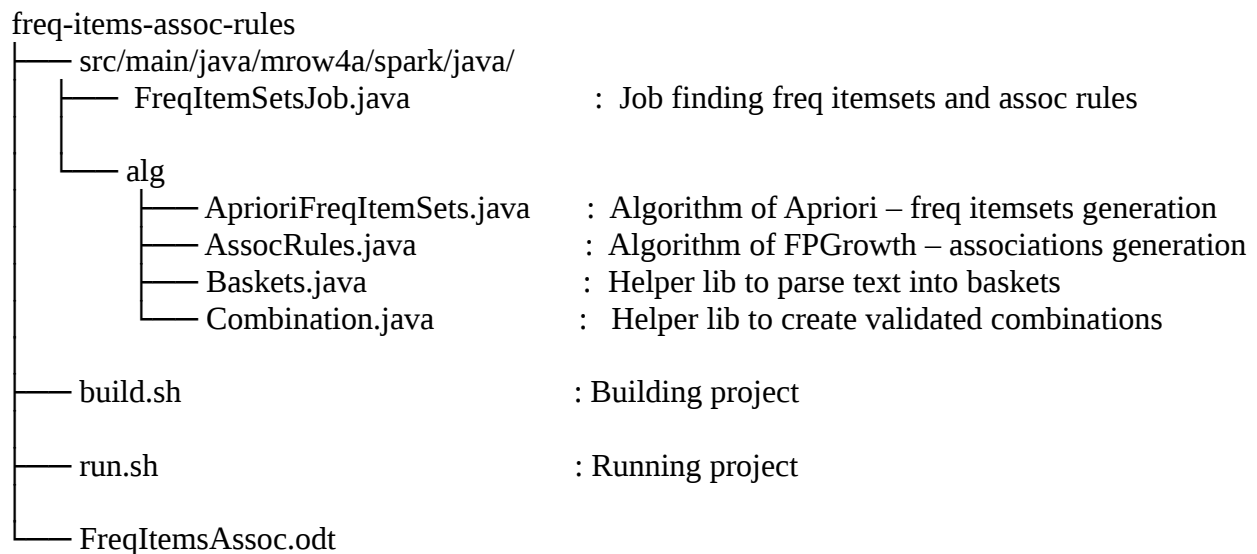# Data Mining 11.10.2017 – Piotr Mrowczynski – Assignment 2

Discovering association rules between itemsets in a sales transaction database (a set of baskets), done by two steps: Finding frequent itemsets with support at least s, Generating association rules with confidence at least c from the itemsets found in the first step.

Project structure

Projects consists of commented code, in which all the steps of preparing dataset, generating combinations, performing Apriori algorithm to find frequent datasets, and based on frequent datasets and their supports, find association rules using FPGrowth algorithm

```
freq-items-assoc-rules
├── src/main/java/mrow4a/spark/java/
│       ├── FreqItemSetsJob.java           :  Job finding freq itemsets and assoc rules
│       └── alg
│               ├── AprioriFreqItemSets.java    :  Algorithm of Apriori – freq itemsets generation
│               ├── AssocRules.java             :  Algorithm of FPGrowth – associations generation
│               ├── Baskets.java                :  Helper lib to parse text into baskets
│               └── Combination.java            :  Helper lib to create validated combinations
├── build.sh                               : Building project
├── run.sh                                 : Running project
└── FreqItemsAssoc.odt
```

Building (Docker required for maven single assembly)

chmod +x freq-items-assoc-rules/build.sh
chmod +x freq-items-assoc-rules/run.sh
freq-items-assoc-rules/build.sh

Running freq items detection and association generation

freq-items-assoc-rules/run.sh

Time: 20749 milliseconds

Dataset: 100000 baskets of items

Association rules detected:

[704] -> [39] with confidence 0.617056856187291
[704] -> [825] with confidence 0.6142697881828316
[39, 704] -> [825] with confidence 0.9349593495934959
[704, 825] -> [39] with confidence 0.9392014519056261
[227] -> [390] with confidence 0.577007700770077
[39, 825] -> [704] with confidence 0.8719460825610783

Frequent Itemsets:

[39, 704, 825]: 1035
[227, 390]: 1049
[789, 829]: 1194
[39, 704]: 1107
[368, 682]: 1193
[217, 346]: 1336
[390, 722]: 1042
[39, 825]: 1187
[368, 829]: 1194
[704, 825]: 1102
[72]: 2852
[322]: 1154
[490]: 1066
[52]: 1983
[513]: 1287
[469]: 1502
[290]: 1793
[632]: 1070
[694]: 2847
[782]: 2767
[843]: 1222
[946]: 1350
[805]: 1789
[227]: 1818
[540]: 1293
[168]: 1538
[21]: 2666
[438]: 4511
[157]: 1140
[373]: 2007
[878]: 2047
[115]: 1775
[546]: 1050
[522]: 2725
[634]: 2492
[274]: 2628
[692]: 4993
[832]: 2062
[874]: 2237
[919]: 3710
[27]: 2165
[964]: 1518
[704]: 1794
[162]: 1450
[676]: 2717
[5]: 1094
[319]: 1371
[641]: 1494
[720]: 3864
[45]: 1728
[403]: 1722
[461]: 1498

[867]: 1530
[294]: 1445
[368]: 7828
[351]: 1641
[377]: 1149
[975]: 1764
[928]: 1034
[449]: 1890
[623]: 1845
[937]: 4681
[982]: 1640
[810]: 1267
[517]: 1201
[285]: 2600
[511]: 1015
[580]: 1667
[229]: 2281
[630]: 1523
[885]: 3043
[357]: 1142
[665]: 1297
[94]: 1201
[10]: 1351
[788]: 2386
[780]: 2306
[500]: 1444
[69]: 2370
[494]: 5102
[296]: 2210
[687]: 1762
[414]: 1160
[78]: 2471
[597]: 2883
[197]: 1230
[348]: 1226
[744]: 2177
[366]: 1031
[614]: 3134
[793]: 3063
[175]: 2791
[809]: 2163
[100]: 1749
[450]: 2082
[210]: 2009
[746]: 1982
[663]: 2354
[54]: 2595
[618]: 1337
[205]: 3605
[201]: 1029
[593]: 2601
[948]: 1149

[58]: 1330
[427]: 1856
[658]: 1881
[834]: 1373
[489]: 3420
[335]: 1345
[784]: 1257
[401]: 3667
[308]: 1402
[722]: 5845
[126]: 1075
[775]: 3771
[140]: 2687
[803]: 2237
[214]: 1893
[906]: 1444
[38]: 2402
[708]: 1090
[887]: 1671
[12]: 3415
[814]: 1672
[591]: 1241
[70]: 2411
[890]: 1437
[120]: 4973
[362]: 4388
[935]: 1742
[429]: 1037
[496]: 1428
[487]: 3135
[526]: 2793
[423]: 1412
[966]: 3921
[661]: 2693
[900]: 1165
[733]: 1141
[25]: 1395
[283]: 4082
[960]: 2732
[504]: 1296
[458]: 1124
[41]: 1353
[913]: 1939
[470]: 4137
[649]: 1292
[472]: 2125
[854]: 2847
[207]: 1214
[85]: 1555
[151]: 2611
[265]: 1359
[32]: 4248

[823]: 1031
[1]: 1535
[825]: 3085
[797]: 2684
[735]: 1689
[766]: 6265
[515]: 1166
[234]: 1416
[111]: 1171
[812]: 1518
[883]: 4902
[405]: 1525
[258]: 1036
[236]: 2618
[944]: 2794
[984]: 1756
[579]: 2164
[571]: 2902
[573]: 1229
[638]: 2288
[548]: 2843
[173]: 1080
[995]: 1521
[90]: 1875
[43]: 1721
[346]: 3470
[829]: 6810
[276]: 2479
[177]: 4629
[326]: 1488
[706]: 1923
[841]: 1927
[605]: 1652
[122]: 1081
[795]: 3361
[740]: 1632
[678]: 1329
[577]: 1695
[674]: 2527
[171]: 1097
[988]: 1164
[991]: 1268
[104]: 1158
[574]: 1297
[381]: 2959
[336]: 1071
[789]: 4309
[530]: 1263
[617]: 2614
[31]: 1666
[181]: 1235
[707]: 1354

[752]: 2578
[183]: 3883
[790]: 1094
[826]: 2022
[992]: 1116
[192]: 2004
[572]: 1589
[774]: 2046
[998]: 2713
[675]: 2976
[334]: 2146
[778]: 2514
[921]: 2425
[378]: 1149
[448]: 1370
[280]: 2108
[154]: 1447
[419]: 5057
[4]: 1394
[266]: 1022
[970]: 2086
[967]: 1695
[129]: 1547
[413]: 2637
[510]: 3281
[480]: 2309
[48]: 2472
[631]: 2793
[800]: 1916
[170]: 1203
[686]: 1495
[815]: 1358
[956]: 3626
[242]: 2325
[736]: 1470
[859]: 1242
[895]: 3385
[947]: 3690
[534]: 1531
[563]: 1065
[758]: 2860
[130]: 1711
[240]: 1399
[93]: 2777
[390]: 2685
[583]: 1389
[55]: 1959
[349]: 2041
[354]: 5835
[57]: 2743
[17]: 1683
[567]: 1102

[684]: 5408
[738]: 2129
[75]: 3151
[914]: 4037
[97]: 1466
[884]: 1645
[279]: 3014
[918]: 3012
[819]: 1257
[682]: 4132
[941]: 1126
[952]: 1574
[147]: 1383
[749]: 1330
[844]: 2814
[594]: 1516
[923]: 1753
[569]: 2835
[125]: 1287
[651]: 1288
[600]: 1192
[527]: 1185
[161]: 2320
[769]: 1622
[860]: 1255
[606]: 2668
[310]: 1390
[132]: 2641
[765]: 1705
[701]: 1283
[871]: 2810
[387]: 2089
[185]: 1529
[963]: 1327
[350]: 3069
[763]: 1862
[116]: 2193
[422]: 1255
[792]: 1306
[950]: 1463
[523]: 2244
[51]: 1612
[486]: 1547
[989]: 1289
[309]: 1262
[718]: 1238
[39]: 4258
[899]: 1252
[820]: 1473
[862]: 3649
[853]: 1804
[509]: 3044

[541]: 3735
[529]: 7057
[673]: 1635
[428]: 1021
[804]: 1315
[198]: 1461
[385]: 1676
[239]: 2742
[259]: 1522
[361]: 1104
[628]: 1102
[145]: 4559
[932]: 1786
[716]: 1199
[561]: 2783
[949]: 1414
[217]: 5375
[846]: 1480
[112]: 2680
[105]: 1100
[521]: 1582
[538]: 3982
[550]: 1203
[35]: 1984
[73]: 2179
[857]: 1588
[71]: 3507
[611]: 1444
[888]: 3686
[598]: 3219
[897]: 1935
[912]: 1009
[196]: 2096
[37]: 1249
[576]: 1337
[163]: 1256
[554]: 1114
[710]: 1044
[6]: 2149
[394]: 1145
[468]: 1089
[8]: 3090
[440]: 1943
[578]: 1290
[978]: 1141
[208]: 1483
[581]: 2943
[798]: 3103
[110]: 1801
[411]: 2047
[424]: 1448
[910]: 1695

[893]: 1947
[204]: 2174
[477]: 2462
[981]: 1542
[639]: 1572
[143]: 1417
[460]: 4438
[471]: 2894
[688]: 1132
[332]: 1861
[886]: 3053
[620]: 2100
[653]: 2634
[68]: 1601
[392]: 2420
[325]: 1022
[28]: 1454
[33]: 1460
[343]: 1599
[516]: 1544
[275]: 1692