Cmpe 493 Introduction to Information Retrieval, Fall 2021 Assignment 1
A Simple Search System for Boolean Queries
Zuhal Didem Aytaç – 2018400045

1. Data Preprocessing

I created a file preprocess.py. This code handles the initial processing of data. It reads the reutx-xxx.sgm files and the stopwords.txt file. It extracts the necessary content from the sgm files. Then with the helper functions, it performs case-fold, punctuation removal and stop-word removal operations on the extracted texts. It keeps the normalized, tokenized text using a dictionary, with new_id as the key.

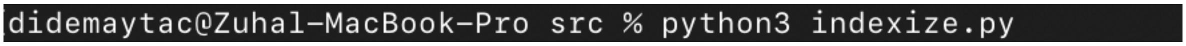
This preprocess module is not run by itself, it is called by the indexing module.

2. The Inverted Index

The indexize module calls the preprocess module described in part 1. The preprocess module returns a dictionary (normalized) with new_id as key and list of tokens as value. The indexize module traverses that dictionary (normalized) and builds a new dictionary (index) with token as keys and list of new_ids as value. This dictionary (index) is a defaultdict, with the default value being set so that no new_ids are duplicated.

After that, the defaultdict (index) is copied into another dictionary (inverted_index) with values being sorted. Finally, the inverted_index dictionary is dumped to a json file (index.json) with token as key and sorted list of new_ids as value.

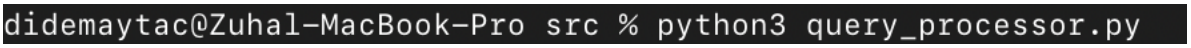
3. The Indexing Module



```
didemaytac@Zuhal-MacBook-Pro src % python3 indexize.py
```

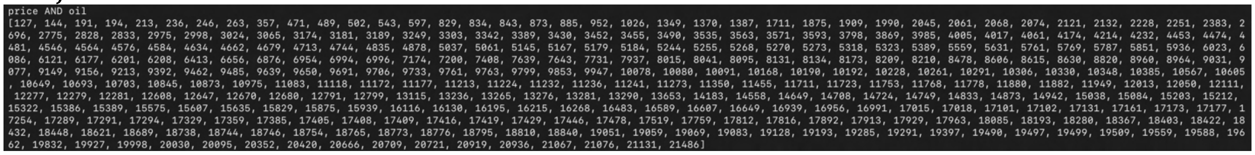
4. Query Processing

You don’t have to specify the query type. Type python3 query_processor.py, then type the query and press enter. The output will be shown at the command line.



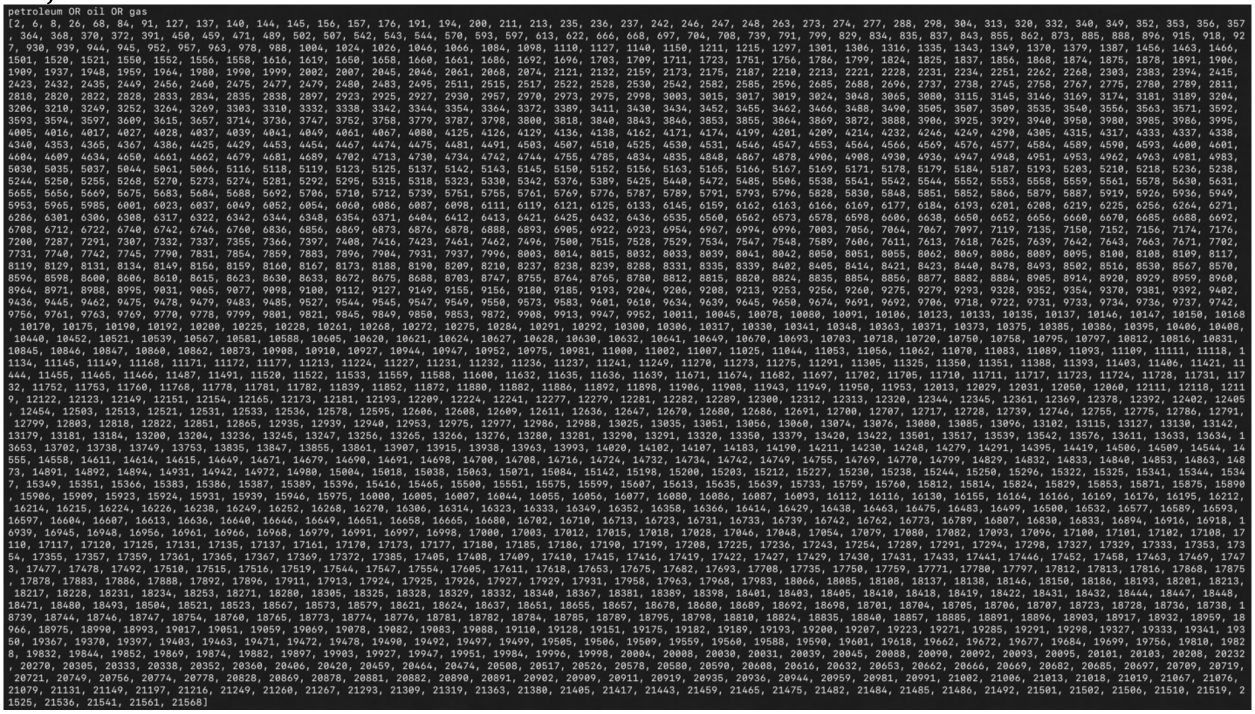
```
didemaytac@Zuhal-MacBook-Pro src % python3 query_processor.py
```

Conjuction:



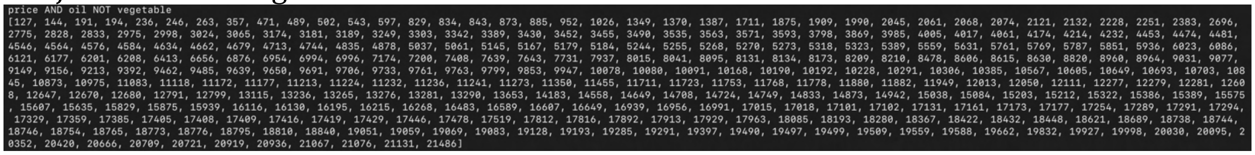
```
price AND oil
[127, 144, 191, 194, 213, 236, 246, 263, 357, 471, 489, 502, 543, 597, 829, 834, 843, 873, 885, 952, 1026, 1349, 1370, 1387, 1711, 1875, 1909, 1990, 2045, 2061, 2068, 2074, 2121, 2132, 2228, 2251, 2383, 2
696, 2775, 2828, 2833, 2975, 2998, 3024, 3065, 3174, 3181, 3189, 3249, 3303, 3342, 3389, 3430, 3452, 3455, 3490, 3535, 3563, 3571, 3593, 3798, 3869, 3985, 4005, 4017, 4061, 4174, 4214, 4232, 4453, 4474, 4
481, 4546, 4564, 4576, 4584, 4634, 4662, 4679, 4713, 4744, 4835, 4878, 5037, 5061, 5145, 5167, 5179, 5184, 5244, 5255, 5268, 5270, 5273, 5318, 5323, 5389, 5559, 5631, 5761, 5769, 5787, 5851, 5936, 6023, 6
086, 6121, 6177, 6201, 6208, 6413, 6656, 6876, 6954, 6994, 6996, 7174, 7200, 7408, 7639, 7643, 7731, 7937, 8015, 8041, 8095, 8131, 8134, 8173, 8209, 8210, 8478, 8606, 8615, 8630, 8820, 8960, 8964, 9031, 9
077, 9149, 9156, 9213, 9392, 9462, 9485, 9639, 9650, 9691, 9706, 9733, 9761, 9763, 9799, 9853, 9947, 10078, 10080, 10091, 10168, 10190, 10192, 10228, 10261, 10291, 10306, 10330, 10348, 10385, 10567, 10605
, 10649, 10693, 10703, 10845, 10873, 10975, 11083, 11118, 11172, 11177, 11213, 11224, 11232, 11236, 11241, 11273, 11350, 11455, 11711, 11723, 11753, 11768, 11778, 11880, 11882, 11949, 12013, 12050, 12111,
 12277, 12279, 12281, 12608, 12647, 12670, 12680, 12791, 12799, 13115, 13236, 13265, 13276, 13281, 13290, 13653, 14183, 14558, 14649, 14708, 14724, 14749, 14833, 14873, 14942, 15038, 15084, 15203, 15212,
15322, 15386, 15389, 15575, 15607, 15635, 15829, 15875, 15939, 16116, 16130, 16195, 16215, 16268, 16483, 16589, 16607, 16649, 16939, 16956, 16991, 17015, 17018, 17101, 17102, 17131, 17161, 17173, 17177, 1
7254, 17289, 17291, 17294, 17329, 17359, 17385, 17405, 17408, 17409, 17416, 17419, 17429, 17446, 17478, 17519, 17759, 17812, 17816, 17892, 17913, 17929, 17963, 18085, 18193, 18280, 18367, 18403, 18422, 18
432, 18448, 18621, 18689, 18738, 18744, 18746, 18754, 18765, 18773, 18776, 18795, 18810, 18840, 19051, 19059, 19069, 19083, 19128, 19193, 19285, 19291, 19397, 19490, 19497, 19499, 19509, 19559, 19588, 196
62, 19832, 19927, 19998, 20030, 20095, 20352, 20420, 20666, 20709, 20721, 20919, 20936, 21067, 21076, 21131, 21486]
```

Disjunction:



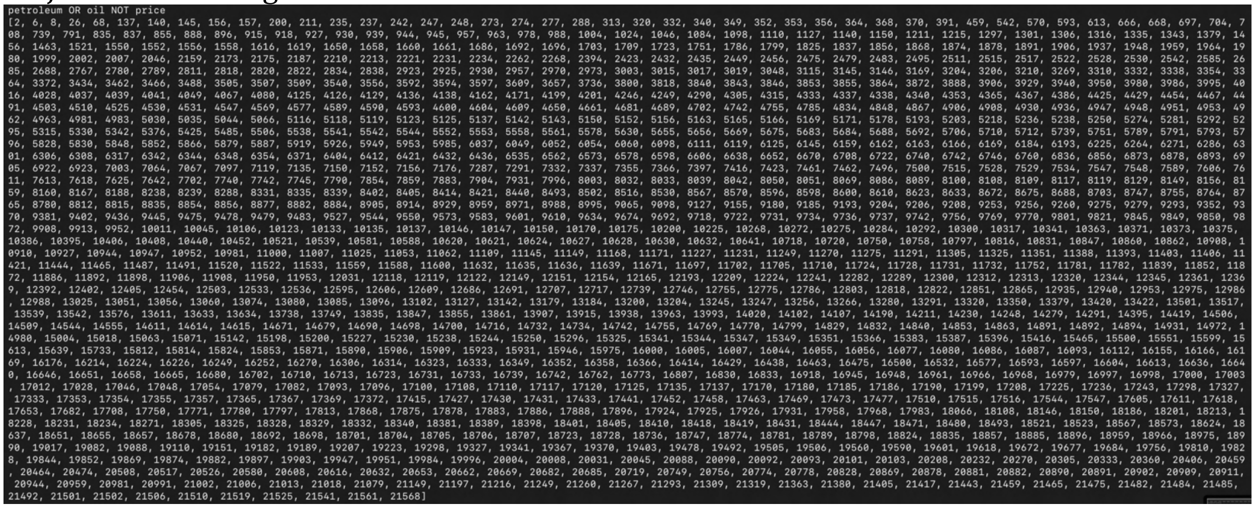
```
petroleum OR oil OR gas
[2, 6, 8, 26, 68, 84, 91, 127, 137, 140, 144, 145, 156, 157, 176, 191, 194, 200, 211, 213, 235, 236, 237, 242, 246, 247, 248, 263, 273, 274, 277, 288, 298, 304, 313, 320, 332, 340, 349, 352, 353, 356, 357
, 364, 368, 370, 372, 391, 450, 459, 471, 489, 502, 507, 542, 543, 544, 570, 593, 597, 613, 622, 666, 668, 697, 704, 708, 739, 791, 799, 829, 834, 835, 837, 843, 855, 862, 873, 885, 888, 896, 915, 918, 92
7, 930, 939, 944, 945, 952, 957, 963, 978, 988, 1004, 1024, 1026, 1046, 1066, 1084, 1098, 1110, 1127, 1140, 1150, 1211, 1215, 1297, 1301, 1306, 1316, 1335, 1343, 1349, 1370, 1379, 1387, 1456, 1463, 1466,
1501, 1520, 1521, 1550, 1552, 1556, 1558, 1616, 1619, 1650, 1658, 1660, 1661, 1686, 1692, 1696, 1703, 1709, 1711, 1723, 1751, 1756, 1786, 1799, 1824, 1825, 1837, 1856, 1868, 1874, 1875, 1878, 1891, 1906,
1909, 1937, 1948, 1959, 1964, 1980, 1990, 1999, 2002, 2007, 2045, 2046, 2061, 2068, 2074, 2121, 2132, 2159, 2173, 2175, 2187, 2210, 2213, 2221, 2228, 2231, 2234, 2251, 2262, 2268, 2303, 2383, 2394, 2415,
2423, 2432, 2435, 2449, 2456, 2460, 2475, 2477, 2479, 2480, 2483, 2495, 2511, 2515, 2517, 2522, 2528, 2530, 2542, 2582, 2585, 2596, 2685, 2688, 2696, 2737, 2738, 2745, 2758, 2767, 2775, 2780, 2789, 2811,
2818, 2820, 2822, 2828, 2833, 2834, 2835, 2838, 2897, 2923, 2925, 2927, 2930, 2957, 2970, 2973, 2975, 2998, 3003, 3015, 3017, 3019, 3024, 3048, 3065, 3080, 3115, 3145, 3146, 3169, 3174, 3181, 3189, 3204,
3206, 3210, 3249, 3252, 3264, 3269, 3303, 3310, 3332, 3338, 3342, 3344, 3354, 3364, 3372, 3389, 3411, 3430, 3434, 3452, 3455, 3462, 3466, 3488, 3490, 3505, 3507, 3509, 3535, 3540, 3556, 3563, 3571, 3592,
3593, 3594, 3597, 3609, 3615, 3657, 3714, 3736, 3747, 3752, 3758, 3779, 3787, 3798, 3800, 3818, 3840, 3843, 3846, 3853, 3855, 3864, 3869, 3872, 3888, 3906, 3925, 3929, 3940, 3950, 3980, 3985, 3986, 3995,
4005, 4016, 4017, 4027, 4028, 4037, 4039, 4041, 4049, 4061, 4067, 4080, 4125, 4126, 4129, 4136, 4138, 4162, 4171, 4174, 4199, 4201, 4209, 4214, 4232, 4246, 4249, 4290, 4305, 4315, 4317, 4333, 4337, 4338,
4340, 4353, 4365, 4367, 4386, 4425, 4429, 4453, 4454, 4467, 4474, 4475, 4481, 4491, 4503, 4507, 4510, 4525, 4530, 4531, 4546, 4547, 4553, 4564, 4566, 4569, 4576, 4577, 4584, 4589, 4590, 4593, 4600, 4601,
4604, 4609, 4634, 4650, 4661, 4662, 4679, 4681, 4689, 4702, 4713, 4730, 4734, 4742, 4744, 4755, 4785, 4834, 4835, 4848, 4867, 4878, 4906, 4908, 4930, 4936, 4947, 4948, 4951, 4953, 4962, 4963, 4981, 4983,
5030, 5035, 5037, 5044, 5061, 5066, 5116, 5118, 5119, 5123, 5125, 5137, 5142, 5143, 5145, 5150, 5152, 5156, 5163, 5165, 5166, 5167, 5169, 5171, 5178, 5179, 5184, 5187, 5193, 5203, 5210, 5218, 5236, 5238,
5244, 5250, 5255, 5268, 5270, 5273, 5274, 5281, 5292, 5295, 5315, 5318, 5323, 5330, 5342, 5376, 5389, 5425, 5440, 5472, 5485, 5506, 5538, 5541, 5542, 5544, 5552, 5553, 5558, 5559, 5561, 5578, 5630, 5631,
5655, 5656, 5669, 5675, 5683, 5684, 5688, 5692, 5706, 5710, 5712, 5739, 5751, 5755, 5761, 5769, 5776, 5787, 5789, 5791, 5793, 5796, 5828, 5830, 5848, 5851, 5852, 5866, 5879, 5887, 5919, 5926, 5936, 5949,
5953, 5965, 5985, 6001, 6023, 6037, 6049, 6052, 6054, 6060, 6086, 6087, 6098, 6111, 6119, 6121, 6125, 6133, 6145, 6159, 6162, 6163, 6166, 6169, 6177, 6184, 6193, 6201, 6208, 6219, 6225, 6256, 6264, 6271,
6286, 6301, 6306, 6308, 6317, 6322, 6342, 6344, 6348, 6354, 6371, 6404, 6412, 6413, 6421, 6425, 6432, 6436, 6535, 6560, 6562, 6573, 6578, 6598, 6606, 6638, 6650, 6652, 6656, 6660, 6670, 6685, 6688, 6692,
6708, 6712, 6722, 6740, 6742, 6746, 6760, 6836, 6856, 6869, 6873, 6876, 6878, 6888, 6893, 6905, 6922, 6923, 6954, 6967, 6994, 6996, 7003, 7056, 7064, 7067, 7097, 7119, 7135, 7150, 7152, 7156, 7174, 7176,
7200, 7287, 7291, 7307, 7332, 7337, 7355, 7366, 7397, 7408, 7416, 7423, 7461, 7462, 7496, 7500, 7515, 7528, 7529, 7534, 7547, 7548, 7589, 7606, 7611, 7613, 7618, 7625, 7639, 7642, 7643, 7663, 7671, 7702,
7731, 7740, 7742, 7745, 7790, 7831, 7854, 7859, 7883, 7896, 7904, 7931, 7937, 7996, 8003, 8014, 8015, 8032, 8033, 8039, 8041, 8042, 8050, 8051, 8055, 8062, 8069, 8086, 8089, 8095, 8100, 8108, 8109, 8117,
8119, 8129, 8131, 8134, 8149, 8156, 8159, 8160, 8167, 8173, 8188, 8190, 8209, 8210, 8237, 8238, 8239, 8288, 8331, 8335, 8339, 8402, 8405, 8414, 8421, 8423, 8440, 8478, 8493, 8502, 8516, 8530, 8567, 8570,
8596, 8598, 8600, 8606, 8610, 8615, 8623, 8630, 8633, 8672, 8675, 8688, 8703, 8747, 8755, 8764, 8765, 8780, 8812, 8815, 8820, 8824, 8835, 8854, 8856, 8877, 8882, 8884, 8905, 8914, 8920, 8929, 8959, 8960,
8964, 8971, 8988, 8995, 9031, 9065, 9077, 9098, 9100, 9112, 9127, 9149, 9155, 9156, 9180, 9185, 9193, 9204, 9206, 9208, 9213, 9253, 9256, 9260, 9275, 9279, 9293, 9328, 9352, 9354, 9370, 9381, 9392, 9402,
9436, 9445, 9462, 9475, 9478, 9479, 9483, 9485, 9527, 9544, 9545, 9547, 9549, 9550, 9573, 9583, 9601, 9610, 9634, 9639, 9645, 9650, 9674, 9691, 9692, 9706, 9718, 9722, 9731, 9733, 9734, 9736, 9737, 9742,
9756, 9761, 9763, 9769, 9770, 9778, 9799, 9801, 9821, 9845, 9849, 9850, 9853, 9872, 9908, 9913, 9947, 9952, 10011, 10045, 10078, 10080, 10091, 10106, 10123, 10133, 10135, 10137, 10146, 10147, 10150, 10168
, 10170, 10175, 10190, 10192, 10200, 10225, 10228, 10261, 10268, 10272, 10275, 10284, 10291, 10292, 10300, 10306, 10317, 10330, 10341, 10348, 10363, 10371, 10373, 10375, 10385, 10386, 10395, 10406, 10408,
 10440, 10452, 10521, 10539, 10567, 10581, 10588, 10605, 10620, 10621, 10624, 10627, 10628, 10630, 10632, 10641, 10649, 10670, 10693, 10703, 10718, 10720, 10750, 10758, 10795, 10797, 10812, 10816, 10831,
10845, 10846, 10847, 10860, 10862, 10873, 10908, 10910, 10927, 10944, 10947, 10952, 10975, 10981, 11000, 11002, 11007, 11025, 11044, 11053, 11056, 11062, 11070, 11083, 11089, 11093, 11109, 11111, 11118, 1
1134, 11145, 11149, 11168, 11171, 11172, 11177, 11213, 11224, 11227, 11231, 11232, 11236, 11237, 11241, 11249, 11270, 11273, 11275, 11291, 11305, 11325, 11350, 11351, 11388, 11393, 11403, 11406, 11421, 11
444, 11455, 11465, 11466, 11487, 11491, 11520, 11522, 11533, 11559, 11588, 11600, 11632, 11635, 11636, 11639, 11671, 11674, 11682, 11697, 11702, 11705, 11710, 11711, 11717, 11723, 11724, 11728, 11731, 117
32, 11752, 11753, 11760, 11768, 11778, 11781, 11782, 11839, 11852, 11872, 11880, 11882, 11886, 11892, 11898, 11906, 11908, 11943, 11949, 11950, 11953, 12013, 12029, 12031, 12050, 12060, 12111, 12118, 1211
9, 12122, 12123, 12149, 12151, 12154, 12165, 12173, 12181, 12193, 12209, 12224, 12241, 12277, 12279, 12281, 12282, 12289, 12300, 12312, 12313, 12320, 12344, 12345, 12361, 12369, 12378, 12392, 12402, 12405
, 12454, 12503, 12513, 12521, 12531, 12533, 12536, 12578, 12595, 12606, 12608, 12609, 12611, 12636, 12647, 12670, 12680, 12686, 12691, 12700, 12707, 12717, 12728, 12739, 12746, 12755, 12775, 12786, 12791,
 12799, 12803, 12818, 12822, 12851, 12865, 12935, 12939, 12940, 12953, 12975, 12977, 12986, 12988, 13025, 13035, 13051, 13056, 13060, 13074, 13076, 13080, 13085, 13096, 13102, 13115, 13127, 13130, 13142,
13179, 13181, 13184, 13200, 13204, 13236, 13245, 13247, 13256, 13265, 13266, 13276, 13280, 13281, 13290, 13291, 13320, 13350, 13379, 13420, 13422, 13501, 13517, 13539, 13542, 13576, 13611, 13633, 13634, 1
3653, 13702, 13738, 13749, 13753, 13835, 13847, 13855, 13861, 13907, 13915, 13938, 13963, 13993, 14020, 14102, 14107, 14183, 14190, 14211, 14230, 14248, 14279, 14291, 14395, 14419, 14506, 14509, 14544, 14
555, 14558, 14611, 14614, 14615, 14649, 14671, 14679, 14690, 14691, 14698, 14700, 14708, 14716, 14724, 14732, 14734, 14742, 14749, 14755, 14769, 14770, 14799, 14829, 14832, 14833, 14840, 14853, 14863, 148
73, 14891, 14892, 14894, 14931, 14942, 14972, 14980, 15004, 15018, 15038, 15063, 15071, 15084, 15142, 15198, 15200, 15203, 15212, 15227, 15230, 15238, 15244, 15250, 15296, 15322, 15325, 15341, 15344, 1534
7, 15349, 15351, 15366, 15383, 15386, 15387, 15389, 15396, 15416, 15465, 15500, 15551, 15575, 15599, 15607, 15613, 15635, 15639, 15733, 15759, 15760, 15812, 15814, 15824, 15829, 15853, 15871, 15875, 15890
, 15906, 15909, 15923, 15924, 15931, 15939, 15946, 15975, 16000, 16005, 16007, 16044, 16055, 16056, 16077, 16080, 16086, 16087, 16093, 16112, 16116, 16130, 16155, 16164, 16166, 16169, 16176, 16195, 16212,
 16214, 16215, 16224, 16226, 16238, 16249, 16252, 16268, 16270, 16306, 16314, 16323, 16333, 16349, 16352, 16358, 16366, 16414, 16429, 16438, 16463, 16475, 16483, 16499, 16500, 16532, 16577, 16589, 16593,
16597, 16604, 16607, 16613, 16636, 16640, 16646, 16649, 16651, 16658, 16665, 16680, 16702, 16710, 16713, 16723, 16731, 16733, 16739, 16742, 16762, 16773, 16789, 16807, 16830, 16833, 16894, 16916, 16918, 1
6939, 16945, 16948, 16956, 16961, 16966, 16968, 16979, 16991, 16997, 16998, 17000, 17003, 17012, 17015, 17018, 17028, 17046, 17048, 17054, 17079, 17080, 17082, 17093, 17096, 17100, 17101, 17102, 17108, 17
110, 17117, 17120, 17125, 17131, 17135, 17137, 17161, 17170, 17173, 17177, 17180, 17185, 17186, 17190, 17199, 17208, 17225, 17236, 17243, 17254, 17289, 17291, 17294, 17298, 17327, 17329, 17333, 17353, 173
54, 17355, 17357, 17359, 17361, 17365, 17367, 17369, 17372, 17385, 17405, 17408, 17409, 17410, 17415, 17416, 17419, 17422, 17427, 17429, 17430, 17431, 17433, 17441, 17446, 17452, 17458, 17463, 17469, 1747
3, 17477, 17478, 17492, 17510, 17515, 17516, 17519, 17544, 17547, 17554, 17605, 17611, 17618, 17653, 17675, 17682, 17693, 17708, 17735, 17750, 17759, 17771, 17780, 17797, 17812, 17813, 17816, 17868, 17875
, 17878, 17883, 17886, 17888, 17892, 17896, 17911, 17913, 17924, 17925, 17926, 17927, 17929, 17931, 17958, 17963, 17968, 17983, 18066, 18085, 18108, 18137, 18138, 18146, 18150, 18186, 18193, 18201, 18213,
 18217, 18228, 18231, 18234, 18253, 18271, 18280, 18305, 18325, 18328, 18329, 18332, 18340, 18367, 18381, 18389, 18398, 18401, 18403, 18405, 18410, 18418, 18419, 18422, 18431, 18432, 18444, 18447, 18448,
18471, 18480, 18493, 18504, 18521, 18523, 18567, 18573, 18579, 18621, 18624, 18637, 18651, 18655, 18657, 18678, 18680, 18689, 18692, 18698, 18701, 18704, 18705, 18706, 18707, 18723, 18728, 18736, 18738, 1
8739, 18744, 18746, 18747, 18754, 18760, 18765, 18773, 18774, 18776, 18781, 18782, 18784, 18785, 18789, 18795, 18798, 18810, 18824, 18835, 18840, 18857, 18885, 18891, 18896, 18903, 18917, 18932, 18959, 18
966, 18975, 18990, 18993, 19017, 19051, 19059, 19069, 19078, 19082, 19083, 19088, 19110, 19128, 19151, 19175, 19182, 19189, 19193, 19200, 19207, 19223, 19271, 19285, 19291, 19298, 19327, 19333, 19341, 193
50, 19367, 19370, 19397, 19403, 19463, 19471, 19472, 19478, 19490, 19492, 19497, 19499, 19505, 19506, 19509, 19559, 19560, 19588, 19590, 19601, 19618, 19662, 19672, 19677, 19684, 19699, 19756, 19810, 1982
8, 19832, 19844, 19852, 19869, 19874, 19882, 19897, 19903, 19927, 19947, 19951, 19984, 19996, 19998, 20004, 20008, 20030, 20031, 20039, 20045, 20088, 20090, 20092, 20093, 20095, 20101, 20103, 20208, 20232
, 20270, 20305, 20333, 20338, 20352, 20360, 20406, 20420, 20459, 20464, 20474, 20508, 20517, 20526, 20578, 20580, 20590, 20608, 20616, 20632, 20653, 20662, 20666, 20669, 20682, 20685, 20697, 20709, 20719,
 20721, 20749, 20756, 20774, 20778, 20828, 20869, 20878, 20881, 20882, 20890, 20891, 20902, 20909, 20911, 20919, 20935, 20936, 20944, 20959, 20981, 20991, 21002, 21006, 21013, 21018, 21019, 21067, 21076,
21079, 21131, 21149, 21197, 21216, 21249, 21260, 21267, 21293, 21309, 21319, 21363, 21380, 21405, 21417, 21443, 21459, 21465, 21475, 21482, 21484, 21485, 21486, 21492, 21501, 21502, 21506, 21510, 21519, 2
1525, 21536, 21541, 21561, 21568]
```

Conjunction and Negation:



```
price AND oil NOT vegetable
[127, 144, 191, 194, 236, 246, 263, 357, 471, 489, 502, 543, 597, 829, 834, 843, 873, 885, 952, 1026, 1349, 1370, 1387, 1711, 1875, 1909, 1990, 2045, 2061, 2068, 2074, 2121, 2132, 2228, 2251, 2383, 2696,
2775, 2828, 2833, 2975, 2998, 3024, 3065, 3174, 3181, 3189, 3249, 3303, 3342, 3389, 3430, 3452, 3455, 3490, 3535, 3563, 3571, 3593, 3798, 3869, 3985, 4005, 4017, 4061, 4174, 4214, 4232, 4453, 4474, 4481,
4546, 4564, 4576, 4584, 4634, 4662, 4679, 4713, 4744, 4835, 4878, 5037, 5061, 5145, 5167, 5179, 5184, 5244, 5255, 5268, 5270, 5273, 5318, 5323, 5389, 5559, 5631, 5761, 5769, 5787, 5851, 5936, 6023, 6086,
6121, 6177, 6201, 6208, 6413, 6656, 6876, 6954, 6994, 6996, 7174, 7200, 7408, 7639, 7643, 7731, 7937, 8015, 8041, 8095, 8131, 8134, 8173, 8209, 8210, 8478, 8606, 8615, 8630, 8820, 8960, 8964, 9031, 9077,
9149, 9156, 9213, 9392, 9462, 9485, 9639, 9650, 9691, 9706, 9733, 9761, 9763, 9799, 9853, 9947, 10078, 10080, 10091, 10168, 10190, 10192, 10228, 10291, 10306, 10385, 10567, 10605, 10649, 10693, 10703, 108
45, 10873, 10975, 11083, 11118, 11172, 11177, 11213, 11224, 11232, 11236, 11241, 11273, 11350, 11455, 11711, 11723, 11753, 11768, 11778, 11880, 11882, 11949, 12013, 12050, 12111, 12277, 12279, 12281, 1260
8, 12647, 12670, 12680, 12791, 12799, 13115, 13236, 13265, 13276, 13281, 13290, 13653, 14183, 14558, 14649, 14708, 14724, 14749, 14833, 14873, 14942, 15038, 15084, 15203, 15212, 15322, 15386, 15389, 15575
, 15607, 15635, 15829, 15875, 15939, 16116, 16130, 16195, 16215, 16268, 16483, 16589, 16607, 16649, 16939, 16956, 16991, 17015, 17018, 17101, 17102, 17131, 17161, 17173, 17177, 17254, 17289, 17291, 17294,
 17329, 17359, 17385, 17405, 17408, 17409, 17416, 17419, 17429, 17446, 17478, 17519, 17812, 17816, 17892, 17913, 17929, 17963, 18085, 18193, 18280, 18367, 18422, 18432, 18448, 18621, 18689, 18738, 18744,
18746, 18754, 18765, 18773, 18776, 18795, 18810, 18840, 19051, 19059, 19069, 19083, 19128, 19193, 19285, 19291, 19397, 19490, 19497, 19499, 19509, 19559, 19588, 19662, 19832, 19927, 19998, 20030, 20095, 2
0352, 20420, 20666, 20709, 20721, 20919, 20936, 21067, 21076, 21131, 21486]
```

Disjunction and Negation:



```
petroleum OR oil NOT price
[2, 6, 8, 26, 68, 137, 140, 145, 156, 157, 200, 211, 235, 237, 242, 247, 248, 273, 274, 277, 288, 313, 320, 332, 340, 349, 352, 353, 356, 364, 368, 370, 391, 459, 542, 570, 593, 613, 666, 668, 697, 704, 7
08, 739, 791, 835, 837, 855, 888, 896, 915, 918, 927, 930, 939, 944, 945, 957, 963, 978, 988, 1004, 1024, 1046, 1084, 1098, 1110, 1127, 1140, 1150, 1211, 1215, 1297, 1301, 1306, 1316, 1335, 1343, 1379, 14
56, 1463, 1521, 1550, 1552, 1556, 1558, 1616, 1619, 1650, 1658, 1660, 1661, 1686, 1692, 1696, 1703, 1709, 1723, 1751, 1786, 1799, 1825, 1837, 1856, 1868, 1874, 1878, 1891, 1906, 1937, 1948, 1959, 1964, 19
80, 1999, 2002, 2007, 2046, 2159, 2173, 2175, 2187, 2210, 2213, 2221, 2231, 2234, 2262, 2268, 2394, 2423, 2432, 2435, 2449, 2456, 2475, 2479, 2483, 2495, 2511, 2515, 2517, 2522, 2528, 2530, 2542, 2585, 26
85, 2688, 2767, 2780, 2789, 2811, 2818, 2820, 2822, 2834, 2838, 2923, 2925, 2930, 2957, 2970, 2973, 3003, 3015, 3017, 3019, 3048, 3115, 3145, 3146, 3169, 3204, 3206, 3210, 3269, 3310, 3332, 3338, 3354, 33
64, 3372, 3434, 3462, 3466, 3488, 3505, 3507, 3509, 3540, 3556, 3592, 3594, 3597, 3609, 3657, 3736, 3800, 3818, 3840, 3843, 3846, 3853, 3855, 3864, 3872, 3888, 3906, 3929, 3940, 3950, 3980, 3986, 3995, 40
16, 4028, 4037, 4039, 4041, 4049, 4067, 4080, 4125, 4126, 4129, 4136, 4138, 4162, 4171, 4199, 4201, 4246, 4249, 4290, 4305, 4315, 4333, 4337, 4338, 4340, 4353, 4365, 4367, 4386, 4425, 4429, 4454, 4467, 44
91, 4503, 4510, 4525, 4530, 4531, 4547, 4569, 4577, 4589, 4590, 4593, 4600, 4604, 4609, 4650, 4661, 4681, 4689, 4702, 4742, 4755, 4785, 4834, 4848, 4867, 4906, 4908, 4930, 4936, 4947, 4948, 4951, 4953, 49
62, 4963, 4981, 4983, 5030, 5035, 5044, 5066, 5116, 5118, 5119, 5123, 5125, 5137, 5142, 5143, 5150, 5152, 5156, 5163, 5165, 5166, 5169, 5171, 5178, 5193, 5203, 5218, 5236, 5238, 5250, 5274, 5281, 5292, 52
95, 5315, 5330, 5342, 5376, 5425, 5485, 5506, 5538, 5541, 5542, 5544, 5552, 5553, 5558, 5561, 5578, 5630, 5655, 5656, 5669, 5675, 5683, 5684, 5688, 5692, 5706, 5710, 5712, 5739, 5751, 5789, 5791, 5793, 57
96, 5828, 5830, 5848, 5852, 5866, 5879, 5887, 5919, 5926, 5949, 5953, 5985, 6037, 6049, 6052, 6054, 6060, 6098, 6111, 6119, 6125, 6145, 6159, 6162, 6163, 6166, 6169, 6184, 6193, 6225, 6264, 6271, 6286, 63
01, 6306, 6308, 6317, 6342, 6344, 6348, 6354, 6371, 6404, 6412, 6421, 6432, 6436, 6535, 6562, 6573, 6578, 6598, 6606, 6638, 6652, 6670, 6708, 6722, 6740, 6742, 6746, 6760, 6836, 6856, 6873, 6878, 6893, 69
05, 6922, 6923, 7003, 7064, 7067, 7097, 7119, 7135, 7150, 7152, 7156, 7176, 7287, 7291, 7332, 7337, 7355, 7366, 7397, 7416, 7423, 7461, 7462, 7496, 7500, 7515, 7528, 7529, 7534, 7547, 7548, 7589, 7606, 76
11, 7613, 7618, 7625, 7642, 7702, 7740, 7742, 7745, 7790, 7854, 7859, 7883, 7904, 7931, 7996, 8003, 8032, 8033, 8039, 8042, 8050, 8051, 8069, 8086, 8089, 8100, 8108, 8109, 8117, 8119, 8129, 8149, 8156, 81
59, 8160, 8167, 8188, 8238, 8239, 8288, 8331, 8335, 8339, 8402, 8405, 8414, 8421, 8440, 8493, 8502, 8516, 8530, 8567, 8570, 8596, 8598, 8600, 8610, 8623, 8633, 8672, 8675, 8688, 8703, 8747, 8755, 8764, 87
65, 8780, 8812, 8815, 8835, 8854, 8856, 8877, 8882, 8884, 8905, 8914, 8929, 8959, 8971, 8988, 8995, 9065, 9098, 9127, 9155, 9180, 9185, 9193, 9204, 9206, 9208, 9253, 9256, 9260, 9275, 9279, 9293, 9352, 93
70, 9381, 9402, 9436, 9445, 9475, 9478, 9479, 9483, 9527, 9544, 9550, 9573, 9583, 9601, 9610, 9634, 9674, 9692, 9718, 9722, 9731, 9734, 9736, 9737, 9742, 9756, 9769, 9770, 9801, 9821, 9845, 9849, 9850, 98
72, 9908, 9913, 9952, 10011, 10045, 10106, 10123, 10133, 10135, 10137, 10146, 10147, 10150, 10170, 10175, 10200, 10225, 10268, 10272, 10275, 10284, 10292, 10300, 10317, 10341, 10363, 10371, 10373, 10375,
10386, 10395, 10406, 10408, 10440, 10452, 10521, 10539, 10581, 10588, 10620, 10621, 10624, 10627, 10628, 10630, 10632, 10641, 10718, 10720, 10750, 10758, 10797, 10816, 10831, 10847, 10860, 10862, 10908, 1
0910, 10927, 10944, 10947, 10952, 10981, 11000, 11007, 11025, 11053, 11062, 11109, 11145, 11149, 11168, 11171, 11227, 11231, 11249, 11270, 11275, 11291, 11305, 11325, 11351, 11388, 11393, 11403, 11406, 11
421, 11444, 11465, 11487, 11491, 11520, 11522, 11533, 11559, 11588, 11600, 11632, 11635, 11636, 11639, 11671, 11697, 11702, 11705, 11710, 11724, 11728, 11731, 11732, 11752, 11781, 11782, 11839, 11852, 118
72, 11886, 11892, 11898, 11906, 11908, 11950, 11953, 12031, 12118, 12119, 12122, 12149, 12151, 12154, 12165, 12193, 12209, 12224, 12241, 12282, 12289, 12300, 12312, 12313, 12320, 12344, 12345, 12361, 1236
9, 12392, 12402, 12405, 12454, 12503, 12533, 12536, 12595, 12606, 12609, 12686, 12691, 12707, 12717, 12739, 12746, 12755, 12775, 12786, 12803, 12818, 12822, 12851, 12865, 12935, 12940, 12953, 12975, 12986
, 12988, 13025, 13051, 13056, 13060, 13074, 13080, 13085, 13096, 13102, 13127, 13142, 13179, 13184, 13200, 13204, 13245, 13247, 13256, 13266, 13280, 13291, 13320, 13350, 13379, 13420, 13422, 13501, 13517,
 13539, 13542, 13576, 13611, 13633, 13634, 13738, 13749, 13835, 13847, 13855, 13861, 13907, 13915, 13938, 13963, 13993, 14020, 14102, 14107, 14190, 14211, 14230, 14248, 14279, 14291, 14395, 14419, 14506,
14509, 14544, 14555, 14611, 14614, 14615, 14671, 14679, 14690, 14698, 14700, 14716, 14732, 14734, 14742, 14755, 14769, 14770, 14799, 14829, 14832, 14840, 14853, 14863, 14891, 14892, 14894, 14931, 14972, 1
4980, 15004, 15018, 15063, 15071, 15142, 15198, 15200, 15227, 15230, 15238, 15244, 15250, 15296, 15325, 15341, 15344, 15347, 15349, 15351, 15366, 15383, 15387, 15396, 15416, 15465, 15500, 15551, 15599, 15
613, 15639, 15733, 15812, 15814, 15824, 15853, 15871, 15890, 15906, 15909, 15923, 15931, 15946, 15975, 16000, 16005, 16007, 16044, 16055, 16056, 16077, 16080, 16086, 16087, 16093, 16112, 16155, 16166, 161
69, 16176, 16214, 16224, 16226, 16249, 16252, 16270, 16306, 16314, 16323, 16333, 16349, 16352, 16358, 16366, 16414, 16429, 16438, 16463, 16475, 16500, 16532, 16577, 16593, 16597, 16604, 16613, 16636, 1664
0, 16646, 16651, 16658, 16665, 16680, 16702, 16710, 16713, 16723, 16731, 16733, 16739, 16742, 16762, 16773, 16807, 16830, 16833, 16918, 16945, 16948, 16961, 16966, 16968, 16979, 16997, 16998, 17000, 17003
, 17012, 17028, 17046, 17048, 17054, 17079, 17082, 17093, 17096, 17100, 17108, 17110, 17117, 17120, 17125, 17135, 17137, 17170, 17180, 17185, 17186, 17190, 17199, 17208, 17225, 17236, 17243, 17298, 17327,
 17333, 17353, 17354, 17355, 17357, 17365, 17367, 17369, 17372, 17415, 17427, 17430, 17431, 17433, 17441, 17452, 17458, 17463, 17469, 17473, 17477, 17510, 17515, 17516, 17544, 17547, 17605, 17611, 17618,
17653, 17682, 17708, 17750, 17771, 17780, 17797, 17813, 17868, 17875, 17878, 17883, 17886, 17888, 17896, 17924, 17925, 17926, 17931, 17958, 17968, 17983, 18066, 18108, 18146, 18150, 18186, 18201, 18213, 1
8228, 18231, 18234, 18271, 18305, 18325, 18328, 18329, 18332, 18340, 18381, 18389, 18398, 18401, 18405, 18410, 18418, 18419, 18431, 18444, 18447, 18471, 18480, 18493, 18521, 18523, 18567, 18573, 18624, 18
637, 18651, 18655, 18657, 18678, 18680, 18692, 18698, 18701, 18704, 18705, 18706, 18707, 18723, 18728, 18736, 18747, 18774, 18781, 18789, 18798, 18824, 18835, 18857, 18885, 18896, 18959, 18966, 18975, 189
90, 19017, 19082, 19088, 19110, 19151, 19182, 19189, 19207, 19223, 19298, 19327, 19341, 19367, 19370, 19403, 19478, 19492, 19505, 19506, 19560, 19590, 19601, 19618, 19672, 19677, 19684, 19756, 19810, 1982
8, 19844, 19852, 19869, 19874, 19882, 19897, 19903, 19947, 19951, 19984, 19996, 20004, 20008, 20031, 20045, 20088, 20090, 20092, 20093, 20101, 20103, 20208, 20232, 20270, 20305, 20333, 20360, 20406, 20459
, 20464, 20474, 20508, 20517, 20526, 20580, 20608, 20616, 20632, 20653, 20662, 20669, 20682, 20685, 20719, 20749, 20756, 20774, 20778, 20828, 20869, 20878, 20881, 20882, 20890, 20891, 20902, 20909, 20911,
 20944, 20959, 20981, 20991, 21002, 21006, 21013, 21018, 21079, 21149, 21197, 21216, 21249, 21260, 21267, 21293, 21309, 21319, 21363, 21380, 21405, 21417, 21443, 21459, 21465, 21475, 21482, 21484, 21485,
21492, 21501, 21502, 21506, 21510, 21519, 21525, 21541, 21561, 21568]
```