

Instructions are below the output

Complete output when the program is run is

-----OUTPUT-----

```
{'radiology': {'doc1': [3, 133, 1246]}, 'lab': {'doc1': [13, 1167, 1713, 1785], 'doc4': [1034], 'doc6': [1721]},  
'stanford': {'doc1': [17, 56, 165, 932, 1030, 1109, 1197, 1789, 1817, 1902, 2240, 2479, 2552], 'doc2': [0,  
42, 81, 109, 173, 454, 579, 675, 2269, 2419, 2439, 2519, 2564, 2848, 2876, 2961, 3299, 3538, 3587],  
'doc3': [519, 618], 'doc4': [1483], 'doc5': [56, 95, 190, 774, 1027], 'doc6': [59, 85, 465, 1303, 1411, 1846,  
2166, 2229, 2249]}, 'university': {'doc1': [26, 2488], 'doc2': [51, 588, 3547], 'doc4': [1492], 'doc5': [65],  
'doc6': [94, 2175, 2238]}, 'school': {'doc1': [37, 65, 1215, 1993, 2125, 2561], 'doc2': [62, 90, 599, 655,  
2582, 3052, 3184, 3596], 'doc5': [76, 104, 792, 1036], 'doc6': [609]}, 'medicine': {'doc1': [47, 75, 1206,  
1225, 1798, 1826, 2003, 2135, 2571], 'doc2': [72, 100, 609, 2573, 2592, 2857, 2885, 3062, 3194, 3606],  
'doc5': [86, 114, 783, 802, 1046]}, 'quantitative': {'doc1': [91, 336, 547, 635]}, 'imaging': {'doc1': [104,  
381, 560, 912, 996]}, 'department': {'doc1': [119], 'doc2': [474, 565, 818, 876, 974, 1439, 1908], 'doc4':  
[0, 794]}, 'search': {'doc1': [143], 'doc2': [151], 'doc4': [1669], 'doc5': [168], 'doc6': [32, 68]}, 'site':  
{ 'doc1': [155, 1369], 'doc2': [163, 2466, 2646], 'doc5': [180, 969], 'doc6': [50, 80]}, 'medical': {'doc1':  
[174, 373, 904, 2175, 2205], 'doc2': [182, 647, 3234, 3264], 'doc5': [199]}, 'sites': {'doc1': [182], 'doc2':  
[190], 'doc5': [207]}, 'ways': {'doc1': [188, 235, 2401], 'doc2': [196, 243, 3460], 'doc5': [213, 260]},  
'person': {'doc1': [208, 255, 2421], 'doc2': [216, 263, 3480], 'doc5': [233, 280]}, 'alumni': {'doc1': [215,  
2193], 'doc2': [223, 463, 1806, 1961, 2459, 3252], 'doc5': [240], 'doc6': [194]}, 'lane': {'doc1': [222, 2200],  
'doc2': [230, 3259], 'doc5': [247]}, 'library': {'doc1': [227, 2213, 2256], 'doc2': [235, 3272, 3315], 'doc5':  
[252], 'doc6': [1576, 1884]}, 'mission': {'doc1': [271]}, 'develop': {'doc1': [282, 580]}, 'apply': {'doc1':  
[294], 'doc3': [3471]}, 'innovative': {'doc1': [300], 'doc4': [873]}, 'techniques': {'doc1': [311, 759]},  
'efficient': {'doc1': [326]}, 'analysis': {'doc1': [349, 622, 892]}, 'display': {'doc1': [362]}, 'data': {'doc1':  
[389, 920]}, 'interdisciplinary': {'doc1': [402]}, 'collaboration': {'doc1': [420], 'doc3': [498, 1674, 3012]},  
'goals': {'doc1': [434]}, 'education': {'doc1': [440, 1319, 1390, 2095, 2183], 'doc2': [3154, 3242], 'doc3':  
[34, 208], 'doc6': [1919]}, 'train': {'doc1': [453]}, 'physicians': {'doc1': [459, 1530]}, 'technologists':  
{ 'doc1': [474]}, 'locally': {'doc1': [488]}, 'worldwide': {'doc1': [500]}, 'latest': {'doc1': [517]},  
'developments': {'doc1': [524]}, 'research': {'doc1': [568, 1329, 1443, 1461, 1481, 1975], 'doc2': [3034],  
'doc4': [884, 1676, 1709]}, 'new': {'doc1': [588, 696, 755, 2348], 'doc2': [3407], 'doc6': [233, 524, 646]},  
'approaches': {'doc1': [592, 738], 'doc4': [1190]}, 'exploration': {'doc1': [610]}, 'assesment': {'doc1':  
[648]}, 'diagnostic': {'doc1': [661, 727]}, 'images': {'doc1': [672]}, 'result': {'doc1': [684], 'doc4': [546]},  
'cost': {'doc1': [712], 'doc3': [2890]}, 'effective': {'doc1': [717], 'doc3': [1224, 1421]}, 'design': {'doc1':  
[778], 'doc4': [76, 741, 924, 1027, 1466, 1559]}, 'planning': {'doc1': [789], 'doc3': [685]}, 'monitoring':  
{ 'doc1': [802], 'doc4': [1621]}, 'therapy': {'doc1': [816]}, 'patient': {'doc1': [824, 1338, 1504]}, 'care':  
{ 'doc1': [832, 1346, 1512, 1843, 1890], 'doc2': [2902, 2949]}, 'deliver': {'doc1': [840]}, 'valid': {'doc1':  
[848]}, 'clinically': {'doc1': [854]}, 'relevant': {'doc1': [865]}, 'visualization': {'doc1': [874]}, 'community':  
{ 'doc1': [941, 2221, 2264, 2315], 'doc2': [709, 3280, 3323, 3374], 'doc3': [284, 958, 1905, 2449, 2608,  
2825, 3002]}, 'locations': {'doc1': [951]}, 'richard': {'doc1': [961]}, 'lucas': {'doc1': [971]}, 'magnetic':
```

{'doc1': [977]}, 'resonance': {'doc1': [986]}, 'center': {'doc1': [1004, 1080, 1895], 'doc2': [1813, 2954], 'doc3': [271, 303, 1854, 2396, 2808, 2844], 'doc6': [1897]}, '1201': {'doc1': [1011]}, 'welch': {'doc1': [1016]}, 'p170': {'doc1': [1025]}, '94305': {'doc1': [1042, 1121], 'doc6': [2261]}, '5488': {'doc1': [1048]}, '650': {'doc1': [1053, 1132], 'doc6': [2267]}, '725': {'doc1': [1057, 1136]}, '8432': {'doc1': [1061]}, 'james': {'doc1': [1066]}, 'clark': {'doc1': [1074]}, '318': {'doc1': [1087]}, 'campus': {'doc1': [1091], 'doc2': [684, 1205, 1590, 1827]}, 'drive': {'doc1': [1098]}, 's344': {'doc1': [1104]}, '5450': {'doc1': [1127]}, '6862': {'doc1': [1140]}, 'directions': {'doc1': [1145, 2446, 2523], 'doc2': [3505]}, '3qd': {'doc1': [1163]}, 'local': {'doc1': [1171]}, 'hotels': {'doc1': [1177], 'doc2': [405]}, 'departments': {'doc1': [1234, 2045], 'doc2': [2601, 3104]}, '3dq': {'doc1': [1256, 1299, 1575, 1781]}, 'laboratory': {'doc1': [1260, 1303], 'doc4': [1002]}, 'navigation': {'doc1': [1271, 1374], 'doc2': [2651], 'doc4': [1578], 'doc5': [835, 974]}, 'section': {'doc1': [1291], 'doc5': [855]}, 'home': {'doc1': [1314, 1385], 'doc5': [868, 873]}, 'industry': {'doc1': [1351, 1625]}, 'overview': {'doc1': [1400, 1452, 1517, 1634, 1848, 1984, 2116, 2231, 2379], 'doc2': [2907, 3043, 3175, 3290, 3438]}, 'visiting': {'doc1': [1409]}, 'fellowships': {'doc1': [1418]}, 'reading': {'doc1': [1430]}, 'list': {'doc1': [1438]}, 'activities': {'doc1': [1470], 'doc2': [345, 1545], 'doc3': [260, 2797]}, 'opportunities': {'doc1': [1490]}, 'patients': {'doc1': [1545], 'doc5': [891]}, 'protocol': {'doc1': [1554]}, 'development': {'doc1': [1563], 'doc3': [414, 1085]}, 'management': {'doc1': [1579]}, 'software': {'doc1': [1590], 'doc6': [383, 1170, 1196, 1735]}, 'case': {'doc1': [1599]}, 'studies': {'doc1': [1604], 'doc5': [43, 155, 344]}, 'testimonials': {'doc1': [1612]}, 'infrastructure': {'doc1': [1643]}, 'services': {'doc1': [1662], 'doc3': [4, 294, 2835], 'doc6': [19, 124, 727, 1063, 1871, 2140]}, 'faculty': {'doc1': [1680, 2078], 'doc2': [1972, 3137], 'doc4': [1653, 1661], 'doc6': [159]}, 'staff': {'doc1': [1692], 'doc3': [938, 1861, 2436], 'doc6': [171]}, 'history': {'doc1': [1698]}, 'resources': {'doc1': [1717, 1807, 2019], 'doc2': [2866, 3078], 'doc6': [2219]}, 'equipment': {'doc1': [1731], 'doc6': [528]}, 'contact': {'doc1': [1741, 2428, 2512], 'doc2': [2471, 3487, 3571], 'doc4': [1793], 'doc5': [998], 'doc6': [1442, 1929]}, 'information': {'doc1': [1749], 'doc2': [2167], 'doc6': [2207]}, 'jobs': {'doc1': [1761], 'doc6': [1835]}, 'getting': {'doc1': [1766, 1835], 'doc2': [2894]}, 'physician': {'doc1': [1864], 'doc2': [2923]}, 'clinical': {'doc1': [1881, 2029], 'doc2': [622, 2940, 3088]}, 'hospital': {'doc1': [1911, 1956], 'doc2': [719, 2970, 3015]}, 'clinics': {'doc1': [1922], 'doc2': [2981]}, 'lucile': {'doc1': [1930], 'doc2': [2989]}, 'packard': {'doc1': [1937], 'doc2': [2996]}, 'children': {'doc1': [1945], 'doc2': [3004]}, 'emergency': {'doc1': [1965], 'doc2': [3024]}, 'news': {'doc1': [2012, 2388], 'doc2': [3071, 3447], 'doc6': [514]}, 'trials': {'doc1': [2038], 'doc2': [3097]}, 'institutes': {'doc1': [2057], 'doc2': [3116]}, 'centers': {'doc1': [2070], 'doc2': [3129], 'doc3': [968, 2459, 2575, 2618], 'doc5': [811]}, 'profiles': {'doc1': [2086], 'doc2': [3145]}, 'training': {'doc1': [2107], 'doc2': [3166], 'doc3': [1008], 'doc6': [498]}, 'programs': {'doc1': [2144], 'doc2': [631, 1486, 3203], 'doc5': [821]}, 'admissions': {'doc1': [2153], 'doc2': [3212], 'doc4': [1718]}, 'continuing': {'doc1': [2164], 'doc2': [3223]}, 'health': {'doc1': [2249], 'doc2': [3308], 'doc4': [1614]}, 'newsletter': {'doc1': [2274, 2468], 'doc2': [3333, 3527]}, 'volunteering': {'doc1': [2285], 'doc2': [3344]}, 'public': {'doc1': [2298], 'doc2': [3357], 'doc6': [417, 1005]}, 'service': {'doc1': [2305], 'doc2': [3364]}, 'partnerships': {'doc1': [2325], 'doc2': [3384]}, 'renewal': {'doc1': [2338], 'doc2': [3397]}, 'building': {'doc1': [2352], 'doc2': [927, 3411], 'doc3': [162, 371]}, 'projects': {'doc1': [2361], 'doc2': [3420]}, 'careers': {'doc1': [2393], 'doc2': [3452]}, 'maps': {'doc1': [2439], 'doc2': [3498]}, 'dean': {'doc1': [2461], 'doc2': [3520]}, 'footer': {'doc1': [2499], 'doc2': [3558], 'doc5': [985], 'doc6': [2096]}, 'links': {'doc1': [2506], 'doc2': [2413, 3565], 'doc4': [1787], 'doc5': [939, 992], 'doc6': [1695, 1855]}, 'members': {'doc1': [2534], 'doc5': [945]}, '2009': {'doc1': [2547], 'doc2': [3582], 'doc5': [1022]}, 'terms': {'doc1': [2580], 'doc2': [3615], 'doc5': [1055], 'doc6': [2280]}, 'use': {'doc1': [2589], 'doc2': [3624], 'doc5':

[1064], 'doc6': [2289]], 'powered': {'doc1': [2593], 'doc2': [3628], 'doc5': [1068]}, 'irt': {'doc1': [2604], 'doc2': [3639], 'doc5': [1079]}, 'anesthesia': {'doc2': [9, 118, 488, 832, 988, 1378, 2428, 2448, 2528, 2613]}, '50th': {'doc2': [20, 129, 427, 518, 1754, 1921, 2330, 2624]}, 'year': {'doc2': [25, 134, 432, 759, 2629], 'doc6': [616]}, 'celebration': {'doc2': [30, 139, 333, 437, 1759, 2335, 2634]}, 'enhanced': {'doc2': [277]}, 'experience': {'doc2': [286], 'doc3': [2640]}, 'free': {'doc2': [305]}, 'flash': {'doc2': [310]}, 'player': {'doc2': [316]}, 'things': {'doc2': [356]}, 'weekend': {'doc2': [369, 1073], 'doc3': [590]}, 'schedule': {'doc2': [377]}, 'scientific': {'doc2': [386, 1244]}, 'program': {'doc2': [397, 1771], 'doc3': [553, 1842, 2194, 2757, 2898, 3122, 3489], 'doc4': [893, 936]}, 'welcome': {'doc2': [412]}, 'dear': {'doc2': [449]}, 'approaching': {'doc2': [502]}, 'anniversary': {'doc2': [523]}, 'founding': {'doc2': [542]}, 'separate': {'doc2': [556]}, 'moved': {'doc2': [662]}, 'palo': {'doc2': [699]}, 'alto': {'doc2': [704]}, 'san': {'doc2': [733]}, 'francisco': {'doc2': [737]}, '1959': {'doc2': [750]}, 'later': {'doc2': [764]}, 'john': {'doc2': [773]}, 'bunker': {'doc2': [778]}, 'named': {'doc2': [789]}, 'chair': {'doc2': [805, 2324]}, 'previously': {'doc2': [843]}, 'division': {'doc2': [856], 'doc6': [2154]}, 'surgery': {'doc2': [890]}, 'celebrate': {'doc2': [901, 1894]}, 'years': {'doc2': [918]}, 'premier': {'doc2': [966]}, 'united': {'doc2': [1006]}, 'states': {'doc2': [1013]}, 'invite': {'doc2': [1023]}, 'guests': {'doc2': [1043]}, 'farm': {'doc2': [1057, 1886]}, 'september': {'doc2': [1084]}, '2010': {'doc2': [1100]}, 'kick': {'doc2': [1113]}, 'friday': {'doc2': [1129], 'doc3': [3251]}, 'evening': {'doc2': [1136, 1701, 2089]}, 'reception': {'doc2': [1144]}, 'rodin': {'doc2': [1161]}, 'sculpture': {'doc2': [1167]}, 'garden': {'doc2': [1177]}, 'cantor': {'doc2': [1188]}, 'museum': {'doc2': [1195]}, 'saturday': {'doc2': [1212, 1495, 1692, 2080], 'doc3': [3275]}, 'morning': {'doc2': [1221]}, 'include': {'doc2': [1234], 'doc6': [624]}, 'session': {'doc2': [1255]}, 'consisting': {'doc2': [1263]}, 'discussions': {'doc2': [1277]}, 'current': {'doc2': [1304, 1370, 1478]}, 'colleagues': {'doc2': [1323, 1936]}, 'provide': {'doc2': [1344]}, 'update': {'doc2': [1360]}, 'related': {'doc2': [1389]}, 'practice': {'doc2': [1397]}, 'reminiscences': {'doc2': [1415]}, 'major': {'doc2': [1452]}, 'accomplishments': {'doc2': [1458]}, 'afternoon': {'doc2': [1504]}, 'left': {'doc2': [1522]}, 'open': {'doc2': [1527], 'doc3': [433]}, 'optional': {'doc2': [1536, 1734]}, 'hikes': {'doc2': [1556]}, 'visits': {'doc2': [1562]}, 'pacific': {'doc2': [1576]}, 'ocean': {'doc2': [1584]}, 'tours': {'doc2': [1597]}, 'just': {'doc2': [1610]}, 'time': {'doc2': [1615, 2286], 'doc4': [754, 1236]}, 'enjoy': {'doc2': [1623]}, 'visit': {'doc2': [1644], 'doc6': [1186]}, 'special': {'doc2': [1650]}, 'places': {'doc2': [1658]}, 'people': {'doc2': [1669]}, 'bay': {'doc2': [1683]}, 'area': {'doc2': [1687]}, 'feature': {'doc2': [1714]}, 'black': {'doc2': [1724], 'doc3': [278, 2819]}, 'tie': {'doc2': [1730]}, 'dinner': {'doc2': [1743]}, 'francis': {'doc2': [1786]}, 'arrillaga': {'doc2': [1796]}, 'sincerely': {'doc2': [1837, 2291]}, 'hope': {'doc2': [1847]}, 'able': {'doc2': [1864]}, 'return': {'doc2': [1872]}, 'friends': {'doc2': [1947]}, 'attendance': {'doc2': [1980]}, 'limited': {'doc2': [1999], 'doc3': [2368]}, '450': {'doc2': [2010]}, 'participants': {'doc2': [2014], 'doc3': [2126, 2906, 3049, 3130]}, 'including': {'doc2': [2027], 'doc6': [1141]}, 'spouse': {'doc2': [2037]}, 'guest': {'doc2': [2044]}, 'constraints': {'doc2': [2061]}, 'location': {'doc2': [2097]}, 'mark': {'doc2': [2116]}, 'calendars': {'doc2': [2126]}, 'register': {'doc2': [2144, 2357]}, 'early': {'doc2': [2153]}, 'need': {'doc2': [2183]}, 'website': {'doc2': [2199, 2383], 'doc6': [1487, 1974]}, 'questions': {'doc2': [2223]}, 'comments': {'doc2': [2236]}, 'reached': {'doc2': [2254]}, 'mhr': {'doc2': [2265]}, 'edu': {'doc2': [2278]}, 'myer': {'doc2': [2301]}, 'mike': {'doc2': [2306]}, 'rosenthal': {'doc2': [2311]}, 'committee': {'doc2': [2347]}, 'registration': {'doc2': [2370, 2400]}, 'fax': {'doc2': [2391]}, 'mail': {'doc2': [2395]}, 'consider': {'doc2': [2489]}, 'making': {'doc2': [2498]}, 'donation': {'doc2': [2507]}, 'aim': {'doc2': [2547]}, 'nav': {'doc2': [2662, 2668, 2674, 2682, 2690, 2698, 2704, 2712, 2720, 2728, 2738, 2748, 2758, 2764, 2772, 2780, 2788, 2794, 2802, 2810, 2818, 2824, 2832, 2840]}, 'a3c': {'doc3': [0]}, 'lead': {'doc3': [13, 91, 149, 358, 892, 989, 2978, 3230, 3484, 3579]}, 'leading': {'doc3': [18, 192]},

'activism': {'doc3': [44, 218]}, 'diversity': {'doc3': [57, 231]}, 'application': {'doc3': [96, 2315, 2327, 3584]}, 'read': {'doc3': [115, 3528], 'doc6': [692, 806, 905, 956]}, 'text': {'doc3': [124, 3537]}, 'carefully': {'doc3': [135]}, 'project': {'doc3': [154, 363, 677, 760, 3379], 'doc6': [1326, 1434]}, 'communities': {'doc3': [171, 380, 537, 1260]}, 'color': {'doc3': [186, 395]}, 'asian': {'doc3': [245, 2782]}, 'american': {'doc3': [251, 2788]}, 'centro': {'doc3': [317, 2858]}, 'chicano': {'doc3': [324, 2865]}, 'proud': {'doc3': [336]}, 'announce': {'doc3': [345]}, 'leadership': {'doc3': [403, 1074, 1169, 1354, 1486, 2085, 2629]}, 'course': {'doc3': [426, 3311]}, 'undergraduate': {'doc3': [441]}, 'students': {'doc3': [455, 627, 790, 1201, 1939, 2029, 2383, 2403, 2549, 3401], 'doc6': [146]}, 'interested': {'doc3': [464]}, 'exploring': {'doc3': [478]}, 'issues': {'doc3': [488]}, 'ethnic': {'doc3': [530, 951]}, 'consists': {'doc3': [561]}, 'fall': {'doc3': [577, 851]}, 'quarter': {'doc3': [582, 782, 843], 'doc6': [1609]}, 'retreat': {'doc3': [598, 1919, 2021, 2677, 3158, 3196, 3243, 3445]}, 'fellow': {'doc3': [611]}, 'continues': {'doc3': [636]}, 'weekly': {'doc3': [651, 2104, 3304]}, 'class': {'doc3': [658, 2111], 'doc6': [1259, 1283, 1352]}, 'meetings': {'doc3': [664, 2117, 3318]}, 'culminates': {'doc3': [698]}, 'student': {'doc3': [716], 'doc4': [1701], 'doc6': [709, 1827]}, 'designed': {'doc3': [724, 1501]}, 'executed': {'doc3': [737]}, 'social': {'doc3': [746, 901, 1051, 1276, 1445, 1463, 1594, 1799, 2067, 2168, 2273, 3365]}, 'change': {'doc3': [753, 908, 1058, 1283, 1452, 1470, 1601, 1806, 2074, 2175, 2280, 3372]}, 'winter': {'doc3': [775, 858]}, 'receive': {'doc3': [803]}, 'unit': {'doc3': [815]}, 'academic': {'doc3': [823], 'doc6': [0, 105, 1033, 1600, 1701, 2121, 2198]}, 'credit': {'doc3': [832]}, 'participation': {'doc3': [875]}, 'model': {'doc3': [915, 1065, 1117, 1290, 1477, 1761, 1955, 2287], 'doc4': [1406]}, 'professional': {'doc3': [925]}, 'offered': {'doc3': [981]}, '1999': {'doc3': [1000]}, 'emerging': {'doc3': [1021]}, 'leaders': {'doc3': [1030]}, 'based': {'doc3': [1038, 1313], 'doc4': [70, 918, 1553]}, 'believe': {'doc3': [1100]}, 'framework': {'doc3': [1125]}, 'collaborative': {'doc3': [1138, 1319]}, 'non': {'doc3': [1152, 1337]}, 'hierarchical': {'doc3': [1156, 1341]}, 'useful': {'doc3': [1185]}, 'tool': {'doc3': [1192]}, 'searching': {'doc3': [1210]}, 'methods': {'doc3': [1234], 'doc4': [256, 1422]}, 'working': {'doc3': [1245]}, 'emphasizes': {'doc3': [1296]}, 'value': {'doc3': [1307, 1896]}, 'means': {'doc3': [1370]}, 'diverse': {'doc3': [1380, 2516]}, 'groups': {'doc3': [1388, 3419]}, 'individuals': {'doc3': [1399]}, 'create': {'doc3': [1414]}, 'movements': {'doc3': [1431]}, 'instill': {'doc3': [1513]}, 'young': {'doc3': [1524]}, 'persons': {'doc3': [1530]}, 'strong': {'doc3': [1540], 'doc4': [811]}, 'sense': {'doc3': [1547]}, 'civic': {'doc3': [1556]}, 'responsibility': {'doc3': [1562]}, 'desire': {'doc3': [1583, 2499]}, 'seven': {'doc3': [1612]}, 'core': {'doc3': [1618, 1891]}, 'values': {'doc3': [1623]}, 'consciousness': {'doc3': [1630]}, 'self': {'doc3': [1647]}, 'congruence': {'doc3': [1652]}, 'commitment': {'doc3': [1663]}, 'common': {'doc3': [1688], 'doc5': [426]}, 'purpose': {'doc3': [1695]}, 'controversy': {'doc3': [1703]}, 'civility': {'doc3': [1720]}, 'citizenship': {'doc3': [1733]}, 'stressed': {'doc3': [1745]}, 'reinforced': {'doc3': [1771]}, 'action': {'doc3': [1789]}, 'creation': {'doc3': [1825]}, 'included': {'doc3': [1872]}, 'eighth': {'doc3': [1884]}, 'orient': {'doc3': [1932]}, 'discussion': {'doc3': [1969]}, 'experiential': {'doc3': [1984]}, 'exercises': {'doc3': [1997]}, 'following': {'doc3': [2007, 3077]}, 'continue': {'doc3': [2043]}, 'learn': {'doc3': [2055], 'doc6': [1358]}, 'plan': {'doc3': [2149]}, 'execute': {'doc3': [2158]}, 'activity': {'doc3': [2182]}, 'small': {'doc3': [2207, 3413]}, 'group': {'doc3': [2213], 'doc5': [377]}, 'setting': {'doc3': [2219]}, 'increasing': {'doc3': [2235]}, 'understanding': {'doc3': [2252]}, 'practical': {'doc3': [2305]}, 'selection': {'doc3': [2341]}, 'process': {'doc3': [2351]}, 'space': {'doc3': [2359], 'doc4': [1589]}, 'selected': {'doc3': [2420]}, 'consideration': {'doc3': [2472]}, 'given': {'doc3': [2486]}, 'representative': {'doc3': [2524]}, 'sample': {'doc3': [2539]}, 'prior': {'doc3': [2583, 3183]}, 'involvement': {'doc3': [2589]}, 'required': {'doc3': [2658]}, 'logistics': {'doc3': [2667]}, 'transportation': {'doc3': [2685]}, 'overnight': {'doc3': [2700]}, 'accommodations': {'doc3': [2710]}, 'meals': {'doc3': [2729]}, 'provided': {'doc3': [2743]},

'funded': {'doc3': [2768]}, 'participant': {'doc3': [2919]}, 'expectations': {'doc3': [2931]}, 'main': {'doc3': [2961]}, 'purposes': {'doc3': [2966]}, 'foster': {'doc3': [2989]}, 'cross': {'doc3': [2996]}, 'requires': {'doc3': [3031]}, 'present': {'doc3': [3065, 3207]}, 'requirements': {'doc3': [3087]}, 'mandatory': {'doc3': [3104]}, 'attend': {'doc3': [3143, 3297]}, 'pre': {'doc3': [3154]}, 'meeting': {'doc3': [3166]}, 'thursday': {'doc3': [3174, 3333]}, 'entire': {'doc3': [3223]}, 'offsite': {'doc3': [3235]}, 'october': {'doc3': [3258, 3284]}, '24th': {'doc3': [3266]}, '25th': {'doc3': [3292]}, '00pm': {'doc3': [3349]}, 'complete': {'doc3': [3354]}, 'determined': {'doc3': [3387]}, 'assigned': {'doc3': [3429]}, 'wish': {'doc3': [3460]}, 'click': {'doc3': [3497]}, 'link': {'doc3': [3510]}, 'doc4': [1099]}, 'like': {'doc3': [3558]}, 'aeronautics': {'doc4': [14, 1136]}, 'astronautics': {'doc4': [28, 1152]}, 'aeroastro': {'doc4': [41, 1529]}, 'logo': {'doc4': [51]}, 'doc6': [2116]}, 'computational': {'doc4': [56, 163, 830, 904, 1176, 1274, 1539]}, 'decades': {'doc4': [103]}, 'giant': {'doc4': [111]}, 'strides': {'doc4': [117]}, 'achieved': {'doc4': [135]}, 'aspects': {'doc4': [152]}, 'aerospace': {'doc4': [177, 844, 982, 1017]}, 'engineering': {'doc4': [187, 676, 854]}, 'higher': {'doc4': [199]}, 'fidelity': {'doc4': [206, 1265]}, 'mathematical': {'doc4': [215]}, 'models': {'doc4': [228]}, 'better': {'doc4': [235]}, 'approximation': {'doc4': [242]}, 'faster': {'doc4': [268]}, 'solution': {'doc4': [275]}, 'algorithms': {'doc4': [284, 1348]}, 'developed': {'doc4': [305]}, 'aerodynamic': {'doc4': [319]}, 'structural': {'doc4': [331, 1603]}, 'aeroacoustic': {'doc4': [342]}, 'aeroelastic': {'doc4': [355]}, 'aerothermal': {'doc4': [367]}, 'control': {'doc4': [383, 1566]}, 'applications': {'doc4': [391]}, 'computing': {'doc4': [417, 509, 992]}, 'doc6': [9, 114, 717, 784, 1042, 2130]}, 'speed': {'doc4': [427]}, 'barriers': {'doc4': [433]}, 'shattered': {'doc4': [452]}, 'hardware': {'doc4': [465]}, 'manufacturers': {'doc4': [474]}, 'parallel': {'doc4': [492]}, 'cluster': {'doc4': [501]}, 'ubiquitous': {'doc4': [530]}, 'numerical': {'doc4': [553]}, 'simulation': {'doc4': [563]}, 'increasingly': {'doc4': [578]}, 'complemented': {'doc4': [591]}, 'cases': {'doc4': [616]}, 'replaced': {'doc4': [622]}, 'physical': {'doc4': [631]}, 'tests': {'doc4': [640]}, 'enhance': {'doc4': [649, 763]}, 'reliability': {'doc4': [661]}, 'designs': {'doc4': [688]}, 'improve': {'doc4': [696]}, 'productivity': {'doc4': [708]}, 'engineers': {'doc4': [724]}, 'reduce': {'doc4': [734]}, 'cycle': {'doc4': [748]}, 'performance': {'doc4': [778]}, 'presence': {'doc4': [818]}, 'carried': {'doc4': [953]}, 'primarily': {'doc4': [965]}, 'frg': {'doc4': [1046]}, 'focuses': {'doc4': [1050]}, 'multidisciplinary': {'doc4': [1061]}, 'frameworks': {'doc4': [1079]}, 'different': {'doc4': [1104]}, 'physics': {'doc4': [1114]}, 'pertaining': {'doc4': [1122]}, 'multiscale': {'doc4': [1165]}, 'deal': {'doc4': [1210]}, 'large': {'doc4': [1220]}, 'ranges': {'doc4': [1226]}, 'spatial': {'doc4': [1245]}, 'scales': {'doc4': [1253]}, 'high': {'doc4': [1260]}, 'schemes': {'doc4': [1288]}, 'enable': {'doc4': [1305]}, 'predictive': {'doc4': [1312]}, 'simulations': {'doc4': [1323]}, 'optimization': {'doc4': [1335]}, 'handle': {'doc4': [1368]}, 'complex': {'doc4': [1375]}, 'integrated': {'doc4': [1383]}, 'systems': {'doc4': [1394, 1595]}, 'reduction': {'doc4': [1412]}, 'integrate': {'doc4': [1439]}, 'computation': {'doc4': [1449]}, '1997': {'doc4': [1473]}, '2012': {'doc4': [1478]}, 'doc6': [604]}, 'rights': {'doc4': [1507]}, 'reserved': {'doc4': [1514]}, 'sustainable': {'doc4': [1632]}, 'aviation': {'doc4': [1644]}, 'labs': {'doc4': [1685]}, 'curriculum': {'doc4': [1690]}, 'financial': {'doc4': [1731]}, 'aid': {'doc4': [1741]}, 'events': {'doc4': [1745]}, 'affiliates': {'doc4': [1752]}, 'membership': {'doc4': [1763]}, 'giving': {'doc4': [1774]}, 'als': {'doc5': [9, 123, 312, 536]}, 'consortium': {'doc5': [13, 127, 297, 316]}, 'epidemiological': {'doc5': [27]}, 'aces': {'doc5': [51, 163, 352, 564, 599, 830, 863]}, 'epidemiologic': {'doc5': [141, 330]}, 'international': {'doc5': [363]}, 'researchers': {'doc5': [386, 917]}, 'specialties': {'doc5': [408]}, 'goal': {'doc5': [433]}, 'identify': {'doc5': [444]}, 'environmental': {'doc5': [453]}, 'lifestyle': {'doc5': [467]}, 'genetic': {'doc5': [481]}, 'risk': {'doc5': [489]}, 'factors': {'doc5': [494]}, 'amyotrophic': {'doc5': [506]}, 'lateral': {'doc5': [518]}, 'sclerosis': {'doc5': [526]}, 'lou': {'doc5': [543]}, 'gehrig': {'doc5': [547]}, 'disease': {'doc5': [556]}, 'director': {'doc5': [569, 607]}, 'lorene': {'doc5':

[578]], 'nelson': {'doc5': [585]}, 'phd': {'doc5': [592, 632]}, 'valerie': {'doc5': [616]}, 'mcguire': {'doc5': [624]}, 'mph': {'doc5': [636, 671]}, 'database': {'doc5': [640]}, 'manager': {'doc5': [649]}, 'barbara': {'doc5': [657]}, 'topol': {'doc5': [665]}, 'web': {'doc5': [675]}, 'page': {'doc5': [679]}, 'doc6': [1205]}, 'coordinator': {'doc5': [684]}, 'jacqueline': {'doc5': [696]}, 'itnyre': {'doc5': [707]}, 'mls': {'doc5': [714]}, 'administrative': {'doc5': [718]}, 'associate': {'doc5': [733]}, 'christine': {'doc5': [743]}, 'dorosin': {'doc5': [753]}, 'families': {'doc5': [904]}, 'additinal': {'doc5': [929]}, 'login': {'doc5': [958, 1016]}, 'doc6': [2062, 2068]}, 'member': {'doc5': [1009]}, 'acs': {'doc6': [28, 1685, 2092]}, 'block': {'doc6': [39, 2103]}, 'visitors': {'doc6': [181]}, 'help': {'doc6': [201, 331, 401, 1822]}, 'network': {'doc6': [206, 1793]}, 'connections': {'doc6': [214, 1801]}, 'computer': {'doc6': [237, 445]}, 'device': {'doc6': [249]}, 'needs': {'doc6': [261]}, 'internet': {'doc6': [267]}, 'access': {'doc6': [276, 354]}, 'multimedia': {'doc6': [283, 1210, 1237, 1771]}, 'studio': {'doc6': [294]}, 'image': {'doc6': [301]}, 'audio': {'doc6': [307]}, 'video': {'doc6': [317]}, 'editing': {'doc6': [323]}, 'classroom': {'doc6': [336]}, 'support': {'doc6': [346]}, 'scheduling': {'doc6': [361]}, 'technology': {'doc6': [372, 1052, 1710]}, 'printing': {'doc6': [392, 1813]}, 'print': {'doc6': [406]}, 'machines': {'doc6': [424]}, 'central': {'doc6': [454]}, 'email': {'doc6': [474]}, 'calendar': {'doc6': [480]}, 'afs': {'doc6': [489]}, 'tech': {'doc6': [493, 682, 1866]}, 'helpsu': {'doc6': [507]}, 'blog': {'doc6': [519]}, 'checkout': {'doc6': [538, 660, 850, 896]}, 'items': {'doc6': [547, 650]}, 'sept': {'doc6': [553]}, '2011': {'doc6': [558, 599]}, 'updated': {'doc6': [569]}, 'inventory': {'doc6': [581]}, 'exciting': {'doc6': [637]}, 'meyer': {'doc6': [676, 1272, 1341, 1570]}, 'desk': {'doc6': [687]}, 'annual': {'doc6': [702]}, 'surveys': {'doc6': [736]}, 'ask': {'doc6': [747]}, 'grads': {'doc6': [751]}, 'undergrads': {'doc6': [761]}, 'experiences': {'doc6': [794]}, 'assu': {'doc6': [816]}, 'supports': {'doc6': [821]}, 'projectors': {'doc6': [835, 881]}, 'joseph': {'doc6': [915]}, 'kautz': {'doc6': [922]}, 'elected': {'doc6': [928]}, 'president': {'doc6': [936]}, 'swallt': {'doc6': [949]}, 'windows': {'doc6': [966, 1094]}, 'mac': {'doc6': [980, 1108]}, 'osx': {'doc6': [984, 1112]}, 'snow': {'doc6': [988, 1116]}, 'leopard': {'doc6': [993, 1121]}, 'computers': {'doc6': [1012]}, 'managed': {'doc6': [1022]}, 'upgraded': {'doc6': [1082]}, 'details': {'doc6': [1133]}, 'lists': {'doc6': [1151]}, 'installed': {'doc6': [1160]}, 'montage': {'doc6': [1221]}, 'cs2c': {'doc6': [1232]}, 'production': {'doc6': [1248, 1378, 1782]}, 'flex': {'doc6': [1278, 1347]}, 'look': {'doc6': [1291]}, 'graphic': {'doc6': [1312, 1420]}, 'novel': {'doc6': [1320, 1428]}, 'digital': {'doc6': [1370]}, 'workflows': {'doc6': [1389]}, 'used': {'doc6': [1399]}, 'let': {'doc6': [1453, 1940]}, 'know': {'doc6': [1460, 1947]}, 'think': {'doc6': [1474, 1961]}, 'having': {'doc6': [1508, 1995]}, 'trouble': {'doc6': [1515, 2002]}, 'finding': {'doc6': [1523, 2010]}, 'submit': {'doc6': [1541, 2028]}, 'comment': {'doc6': [1550, 2037]}, 'question': {'doc6': [1561, 2048]}, '2nd': {'doc6': [1584]}, 'floor': {'doc6': [1588]}, 'hours': {'doc6': [1594]}, 'mon': {'doc6': [1617]}, 'thu': {'doc6': [1621]}, '9am': {'doc6': [1625, 1645]}, 'midnight': {'doc6': [1632, 1676]}, 'fri': {'doc6': [1641]}, '5pm': {'doc6': [1649, 1661]}, 'sat': {'doc6': [1653]}, '1pm': {'doc6': [1657, 1669]}, 'sun': {'doc6': [1665]}, 'quick': {'doc6': [1689]}, 'available': {'doc6': [1725]}, 'coursework': {'doc6': [1744]}, 'learning': {'doc6': [1755]}, 'spaces': {'doc6': [1764]}, 'info': {'doc6': [1861, 1892]}, 'residential': {'doc6': [1907]}, 'user': {'doc6': [2057]}, 'sunetid': {'doc6': [2084]}, 'sulair': {'doc6': [2109]}, 'libraries': {'doc6': [2186]}, '723': {'doc6': [2271]}, '2300': {'doc6': [2275]}, 'copyright': {'doc6': [2293]}, 'complaints': {'doc6': [2303]}}

TOTAL No. of Dict Terms Generated ==> 753

['lab', 'project and lead and american', 'radiology center', 'mike or batch', 'computing and mike']

Search for 'lab': {'doc1', 'doc4', 'doc6'}

Search for 'project and lead and american': {'doc3'}

Search for 'radiology center': {'doc1'}

Search for 'mike or batch': {'doc2'}

Search for 'computing and mike': 'No matching Docs' []

-----**OUTPUT END**-----

Instructions to run the application.

- ➔ Download all the zip parts of file. (two parts)
- ➔ Select part1(001) and right click and select “extract here” option.
- ➔ In the folder you will find 4 files. Dataset folder, inverted_index.py, inverted_index.exe, writeup doc.

From command prompt navigate to the extracted folder and run the exe file.

The executable can be run as follows:

C:\Users\sd186076\Desktop\IR_Assignment\dist>inverted_index.exe --help

usage: inverted_index.exe [-h] [-l LIBRARY] [-o OUTPUT] [-q SEARCHQUERY]

[-d INVERTEDINDEX]

...(truncated) this gives the help of the application command.

Running the application.

For all the commands the generic output is : The total number of dictionary items and searches through few sample queries and returns the found docs .

In addition to basic output additional features output can be used as follows.

Command1 (runs the basic command) – Basic command

inverted_index.exe

Command2 (runs the output command) – data from matched documents is printed

inverted_index.exe -o 1

Command3 (runs the dictionary index command) – Complete inverted positional index is printed

inverted_index.exe -d 1

Command4 (runs the nltk command) – Uses nltk libraries stemmer, lemmatizes and tokenizer to build inverted index

```
# inverted_index.exe -l 1
```

Command5 (runs the user query command) – User can pass “query” to search from docs.

```
# inverted_index.exe -q “user query”
```

Note: The above commands can be given individually or in various combinations. Accordingly, the output varies.

```
#inverted_index.exe -l 1 -d 1 -o 1 -q “mike and srinaveen”
```

Execution using the python file (all the above rules apply)

```
Python inverted_index.py -q “mike and srinaveen”
```

SAMPLE OUTPUT FOR USER BOOLEAN QUERIES

OUTPUT1

```
inverted_index.exe -q "navigation"
```

...(truncated output)

TOTAL No. of Dict Terms Generated ==> 753

```
['lab', 'project and lead and american', 'radiology center', 'mike or batch', 'computing and mike', 'navigation']
```

Search for 'lab': {'doc4', 'doc1', 'doc6'}

Search for 'project and lead and american': {'doc3'}

Search for 'radiology center': {'doc1'}

Search for 'mike or batch': {'doc2'}

Search for 'computing and mike': 'No matching Docs' []

```
Search for 'navigation': {'doc4', 'doc2', 'doc1', 'doc5'}
```

OUTPUT2

```
inverted_index.exe -q "navigation and mike"
```


...(output truncated)

Search for 'radiology center': {'doc1'}

Search for 'mike or batch': {'doc2'}

Search for 'computing and mike': 'No matching Docs' []

Search for 'navigation and mike': {'doc2'}

OUTPUT3

inverted_index.exe -q "james and mike"

...(output truncated)

Search for 'mike or batch': {'doc2'}

Search for 'computing and mike': 'No matching Docs' []

Search for 'james and mike': 'No matching Docs' []

Summary of Code

The entire code module has two approaches

- 1) Using NLTK library to tokenize , stemming and lemmatizing (selected by user)
or
- 2) Do the normalization through code logic (default behavior)

Approach 1

Using NLTK library

The code is first extracted from doc and sent as string of text to the function **preprocessor_lib(text)**

The function users stemmer, lemmatizer , stopwords and does casefolding of text sent.

After all the operations are done the words are built back to text and sent to **preprocessor_manual(text)**

Approach 2

The code is first extracted from doc and sent as string of text to the function **preprocessor_manual(text)**

The function parses through each character in text and checks if is alphabet or number.

Then the words normalization is done by casefolding

If any word is less than minimum length of 3 it is removed from text.

if any stopwords exist in the text they are removed.

Each word/number is identified and their locations are returned as tuples.

inverted_index(text) This function process the tuples and constructs an index of the form

{term1: [pos1, pos2, pos3...],

inverted_index_add(inverted, doc_id, doc_index): This function merges all the words from all documents to one dictionary of the form

{term1: {doc1: [pos1, pos2, pos3...], doc2: [pos1, pos2, pos3...] },.... }

search(inverted, query): This function searched through all the keys in dictionary and does an AND operation on the docs of query terms and returns matching docs.

main(): This function does all the steps mentioned in the order of their appearance. The function also finally outputs the data matching from the docs if user selects the “-o” option in command line arguments.

Assumptions:

The docs are only parsed for alphanumeric characters

The search query is for exclusive AND operation only.

The user if selects NLTK library and wishes to see matching output, the returned output is of normalized form.

Predefined stopwords have been used .

Python 3.6 has been used. System support for NLTK library is required.

Runs on windows platform only (tested on multiple windows platforms)