

ANALYSIS

1.

- The time taken by inverted index dictionary is quite much. But process the queries fastly. There are lot of documents retrieved by inverted index as it doesn't take position of query words into account.

X AND Y:

Query : *'place and trees'*

```
Entered query: place and trees
Total documents retrieved: 37
Total comparisons: 1761
Documents ID's: [11394, 9608, 17160, 12683, 11789, 5138, 10008, 14329, 8100, 11684, 5158, 17190, 5802, 12203, 4141, 4654, 7854, 11186, 5178, 10300, 9922, 10704, 11093, 11482, 11483, 16858, 5727, 5087, 4583, 11112, 12008, 17511, 4591, 12656, 11764, 5877, 4473]
```

X OR Y:

Query: *'japan or india'*

```
Entered query: japan or india
Total documents retrieved: 157
Total comparisons: 129
Documents ID's: {9216, 7689, 15881, 12308, 9752, 9753, 10266, 5147, 6176, 16419, 2613, 2614, 7226, 3134, 7743, 9792, 578, 17475, 7238, 10310, 14918, 15945, 7250, 2644, 9817, 6234, 91, 15451, 15457, 11365, 7278, 2671, 14447, 17012, 119, 4221, 15490, 643, 2180, 15500, 19093, 6806, 2202, 4765, 2206, 7839, 7333, 168, 9897, 188, 9408, 7362, 203, 4299, 5328, 9439, 16611, 14564, 10982, 9447, 7678, 7920, 10992, 1778, 9478, 10502, 7436, 7442, 2339, 10532, 19747, 2352, 14137, 7488, 1863, 11079, 7504, 850, 2900, 6998, 8540, 2916, 7013, 15724, 15730, 2939, 16765, 12683, 15251, 9621, 932, 10664, 2985, 2986, 939, 10670, 6063, 3504, 2993, 2994, 7602, 2485, 2491, 15822, 7119, 16848, 5074, 1499, 9696, 2017, 14816, 7145, 3562, 14326, 17406}
```

X AND NOT Y:

Query: *'lightly and not crippled'*

```
Entered query: lightly and not crippled
Total documents retrieved: 29
Total comparisons: 19728
Documents ID's: [10245, 4490, 3980, 9228, 17297, 9622, 19610, 16798, 6817, 19745, 17446, 6825, 9899, 11319, 9144, 5433, 6337, 7752, 7253, 9435, 19167, 15589, 17895, 9577, 2671, 18289, 11767, 15864, 7551]
```

X OR NOT Y:

Query : *'india or not japan'*

```
Time to get posting list of all query words: 0.001995563507080078
Total Number of documents retrieved: 19871
Time to perform all operations: 0.015243291854858398
Documents ID's retrieved: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142]
```

Extended query with AND and OR:

Query : *'rocks and corpora or trees'*

```
Entered query: rocks and corpora or trees
Total documents retrieved: 161
Total comparisons: 237
Documents ID's: [5632, 8193, 10759, 8712, 7691, 16907, 11789, 6158, 5138, 12820, 4125, 5151, 5158, 5159, 4141, 46
54, 9262, 2100, 17974, 5178, 10300, 7233, 8264, 18506, 3147, 4177, 2131, 8790, 12379, 4188, 5727, 612, 4197, 1648
6, 17511, 14955, 11380, 15476, 17524, 4733, 8833, 11394, 5763, 4232, 5768, 12939, 7820, 6798, 18065, 7829, 11934,
5802, 16554, 7854, 6320, 7347, 16574, 9922, 12492, 5841, 4307, 5843, 7894, 11482, 11483, 9951, 12008, 12012, 9966,
8943, 10992, 5877, 4344, 18680, 18684, 6910, 5887, 2816, 4864, 16646, 17158, 17160, 15113, 17163, 12049, 11026, 16
145, 8982, 10008, 2331, 12573, 17190, 7464, 17706, 7480, 7487, 9025, 10569, 17745, 11093, 12632, 7003, 12124, 1212
8, 8034, 4454, 11112, 5997, 12656, 9075, 4473, 10108, 14721, 6020, 9608, 11146, 12683, 1420, 9614, 4511, 12703, 81
00, 6565, 10661, 11684, 12203, 12204, 17323, 11186, 11698, 5561, 6075, 957, 5056, 7105, 7106, 15300, 7112, 10704,
2007, 6104, 5593, 7639, 12763, 16858, 15837, 17367, 5087, 9187, 4068, 18918, 4583, 7655, 7662, 4591, 11764, 10229,
4599, 5624, 14329, 7162]
```

Extended query with AND NOT and OR not:

Query: *'rotate and not determination or not india'*

```
Entered query: rotate and not determination or not india
Total documents retrieved: 19957
Total comparisons: 19660
Documents ID's: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25,
27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53,
55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81,
83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107,
109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130
1, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152,
154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175
6, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197,
199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220
1, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242,
245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266
7, 268, 269, 270, 271, 272, 273, 274, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289,
291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312]
```

As we can see from above outputs that number of comparisons are more than number of document retrieved, due to the reason that some documents are discarded and some are kept according to specified condition.

Observations:

- The **'OR NOT'** operation is most expensive as we have to search in every document.
- The **OR** operation is less expensive as compared to **'OR NOT'**.
- Efficient comparisons are between **'AND'** and **'AND NOT'**.

2.

- The query for this part is the string or a collection of words known as phrase query. In this more relevant documents are retrieved compared to previous.
- The time complexity of positional indexing is more as compared to inverted index but query processing time is less.

Outputs:

1. From com.graphics file:

Query: *'NAVY SCIENTIFIC VISUALISATION'*

```
173995
enter the query : NAVY SCIENTIFIC VISUALIZATION
NAVY SCIENTIFIC VISUALIZATION
['navy', 'scientific', 'visualization']
['37261']
```

2. From rec.motorcycles files:

Query: *'It turns out that'*

```
Scanning rec.motorcycles
173995
enter the query : It turns out that
It turns out that
['it', 'turns', 'out', 'that']
['39638', '39078', '84554', '84433']
```

3. Error Query if length is more than 5:

Query: *'The idea is to clip one polygon'*

```
enter the query : The idea is to clip one polygon
The idea is to clip one polygon
['the', 'idea', 'is', 'to', 'clip', 'one', 'polygon']
query length out of bound,length should be less than 5!!
```

From the last output we can see that it only work for the query having words less than 5.

Observations:

- If we increase the size of string it's not always true that the retrieval of number of documents will also increase. There might be a case that a whole query string belongs to the same document.
- Thus in result, it shows that size of string is not proportional to the time to fetch the documents.