

Analysis

- Inverted Indices provide a fast retrieval method for Boolean queries but can't handle phrase queries.
- The number of comparisons in the merge algorithm can be reduced by using the frequency count of each token and at every run using that count to pick the 2 smallest postings at each run.
- Inverted Indices do not take into account the semantics of the query, the ordering of the words in the documents, etc.
- Users generally give phrase queries rather than queries in the Boolean format.
- Positional indices provide support for phrase queries and proximity search.
- The cons of using positional index is the large size of the index.

Sample Output for Q1:

Query: ultimate and paperback

```
Documents retrieved: [[1, 901]]
Number of documents retrieved: 2
Number of comparisons made: 127
```

Sample Output for Q2:

Query: I think crutches

```
Number of documents retrieved: 1
Documents retrieved: [1011]
```