1. We have indexed 24625 documents since there are 30768 duplicated documents.

2. There are 39858 unique tokens.

3. The total size of the index on disk is 1310.808 KB

4. The structure of the index is {word: posting} and posting is defined as posting(word, freq, pos) where freq is a dictionary that ID is key & freq is value; pos is also a dictionary that ID is key & list of position is value. For example: **skull: 4 -> ID/freq: {12407: 1, 14040: 1, 19331: 1, 21398: 2}, ID/pos: {12407: [2], 14040: [2], 19331: [4], 21398: [4, 7]}**