For the first iteration, the computing density is 0 (there is no actual positive).

The blocking rules used is either the edit distance of the names of the books is larger than 8, or the length of the largest common substring of book names is smaller than the half length of the larger book name.

For example, the edit distance between “Computer Science: A Structured Programming Approach Using C” and “Computer Science: A Structured Prog Approach Using C” is 7 so that it is not blocked.

“Computer Science: A Structured Programming Approach Using C” and “A Structured Programming Approach Using C” has common string “A Structured Programming Approach Using C” which is larger than half the length of “Computer Science: A Structured Programming Approach Using C” so that it is also not blocked.

After that, the set size is 2970 and the computing density is 0.034.

Many alternative blocking rules have been tried (e.g., block all the matchings follows one of the two rules) but they will block many true positive. The reason is that there are many books that are not matched but very similar. For example, “Computer Science: A Structured Programming Approach Using C” and “Computer Science: A Structured Programming Approach Using Java” are very similar but are different. Because the length of our prediction set is about 100, it is very hard to find a blocking rule that has higher density but also don’t block true positives.