2.3 RAT

1. Given the incidence vectors for Antony, Cleopatra, and Calpurnia, i.e.

   ```
   Antony:    110001
   Cleopatra: 100000
   Calpurnia: 010000
   ```

   what is the incidence vector corresponding to the query “(Antony or Calpurnia) and not Cleopatra”?

2. If we have a corpus of 1 million documents, each of length 2,000 words, and a total vocabulary size of 400,000, what is the approximate maximum size of the postings and the size of the (non-sparse) co-occurrence matrix (which contains a 1 in row *i* and column *j* if word *i* occurs in document *j* and a 0 otherwise), respectively?

3. If the length of two postings lists are **`x`** and **`y`**, then what is the tightest upper bound on the running time of merging the postings lists in an OR query in this manner?

   ```
   A)  O(max{x,y})
   B)  O(min{x,y})
   C)  O(x+y)
   D)  O(xy)
   ```

4. Given the postings list for the word "youth":

   ```
   3: 7, 18, 33, 72, 86, 231;
   5: 17, 191, 291, 430, 432;
   6: 3, 145, 149;
   9: 363, 397;
   ```

   Which of documents 3, 5, 6, and 9 could contain “youth without youth”?