(a) General overview of the system with a small user guide

You will be asked to choose the mode of the output by entering “output=full”, “output=brief” or pressing the enter key. The default output mode is brief and once you choose the other mode, the default mode will change and you can change it in the next query search. Then you need to enter some queries separated by some empty spaces. The keywords you can use are “subj:”, “body: ”, “date< or <= or > or >= or :”, “to: or from: or cc: or bcc: plus email address” and single terms. The program will return all matching row ids and subjects if you chose the brief mode and will return all row ids and the whole email records if you chose the full mode. If you want to quit the program, input nothing and press enter when it asks you to input the query.

(b) Description of the algorithm for efficiently evaluating the queries

Check the input format first to skip unnecessary searches. When there exist multiple queries, check them one by one and when one query cannot return any matches, stop the search and return a negative result to skip the following searches.When processing the single query, finding the ids of the user input in the corresponding database and return the row id so that it is easy to find the subject and the whole records.

(c) Testing strategy

Using the print function, we ensure that the correct values are being returned. We also print the type of variable to test if the data type is suitable for generating the output. Then, before being put together, each function is tested separately on inputs of different sizes (10 and 1k). Then we combine all the functions to our program and test with a large size (1k). On the other hand, we tried to input some incorrect input data like gibberish to ensure no result can be returned from the program. We also compare the test output with our generated output by using Linux commands to determine whether there is any difference or not.

(d) Group work-breakdown strategy

Gary was responsible for creating terms.txt, query to return records from subject field, query to return records from body field and query to return records from “from” field

Work item	Progress made	Time spent
terms.txt	100%	4hrs
Terms in subject	100%	1hr
Terms in body	100%	1hr

Terms in from	100%	1hr

Xiutong was responsible for creating the data files dates.txt and recs.txt and the queries to return records from terms found in “to”, “cc”, “bcc” respectively.

Work item	Progress made	Time spent
dates.txt	100%	3hrs
recs.txt	100%	1hr
Terms in to	100%	1hr
Terms in cc	100%	1hr
Terms in bcc	100%	1hr

Xichen was responsible for creating the data file emails.txt, phase 2, function to process input and queries to return records from date field and then combining all functions together.

Work item	Progress made	Time spent
emails.txt	100%	2h
Phase 2	100%	1h
Input processing	100%	4h
Terms in date	100%	30min
Combining all functions	100%	4h

Method of coordination:
Any problem faced by one of the group members is communicated using the messaging app Wechat. To prevent any coordination problems, all three of us would work on our individual work tasks on lab machines in the general access lab. Hence, if any of us encounter problems, we could easily communicate and help each other.