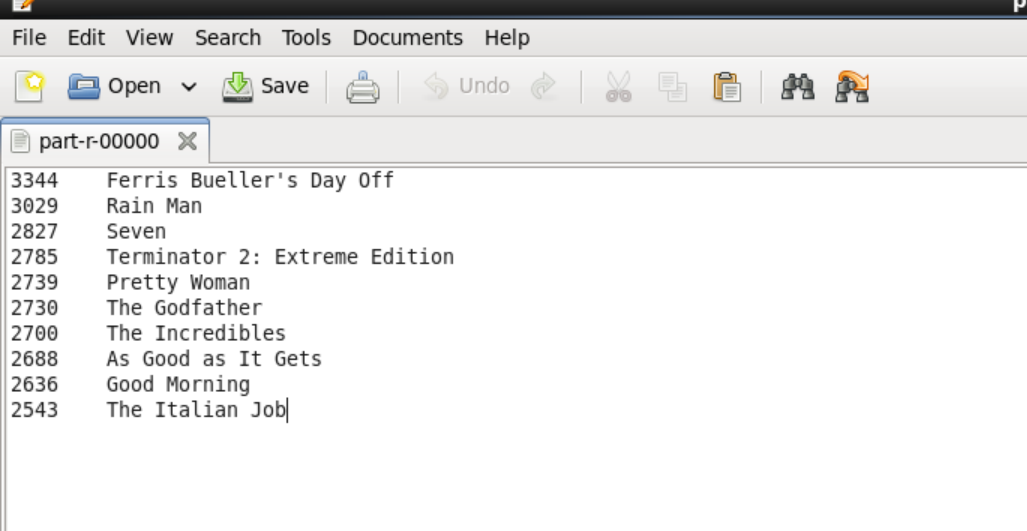
Hw 8

1. (a)



(b)

There are 9494 users have an average rate that more than 4, so the fraction of overly enthusiastic raters is 9494/27555 = 0.3445, and there are 1304 users have an average rate that less than 2, so the fraction of overly pessimistic raters is 1304/27555 = 0.047

To cope with the overly enthusiastic raters and the overly pessimistic raters, so I think the average rating mapreduce job should be done first to find out the user id of this kinds of users and remove them from the processed data according to the id before the run the top-N program.