Answer 4(b)

Now we have to estimate the size of the search engine's indexed corpus.

Frequency (a,b,c) = (frequency(a) * frequency (b) * frequency(c))/ N^ 2 ---------(1)

First Query: **"Time heals everything"**

Let find the N, which is the number of documents in the collections

Frequency (Time heals everything) => it is the estimate size of the result set.

Frequency (Time), Frequency (heals), Frequency (everything) are the number of documents that terms "Time", "heals" and "everything" occur in

N is the Number of documents in the collections

Frequency (Time heals everything) => 10,900,000

Frequency (Time) => 3,180,000,000
Frequency (heals) => 10,600,000
Frequency (everything) => 190,000,000

By putting the above values in equation (1) we get the N

N^2 = (3,180,000,000 * 10,600,000 * 190,000,000) / 10,900,000

N = 766,531,566 ----------(2)

First Query: "Death before dishonor"

Let find the N, which is the number of documents in the collections

Frequency (Death before dishonor) => it is the estimate size of the result set.

Frequency (Death), Frequency (before), Frequency (dishonor) are the number of documents that terms "Death", "before" and "dishonor" occur in

N is the Number of documents in the collections

Frequency (Death before dishonor) => 664,000

Frequency (Death) => 315,000,000
Frequency (before) => 407,000,000
Frequency (dishonor) => 2,204,000

By putting the above values in equation (1) we get the N

N^2 = (315,000,000 * 407,000,000 * 2,204,000)/ 664,000

N = 652,340,341 ----------(3)

After comparing (2) and (3) we can estimate that the corpus size lies nearby the range of 600,000,000 to 800,000,000