EXERCISE 3 - MATHEMATICAL FUNCTIONS

Given the following four documents with the words in them which is the total set of documents we have:

D1: computer computer math computer computer science

D2: physics physics math math physics computer math

D3: physics math biology math science science biology computer biology

D4: physics physics physics physics math computer math computer

***Questions:***

1. **Across all four documents (looking at all the words in all of the documents) show the probability of computer, science, math, physics and biology (can leave in fraction form) occurring**
2. **What is the probability that each word occurs in a document (what per cent of the documents does each word occur in)**
3. **What is average information for the words given probabilities in 1. Above. Three words are 8/30 and 2 words are 3/30.**
4. **If all five words the words occurred equal likely instead of the above – what would the average information be and compare it to the average information from 3. above**

1. **Can you calculate average information for 2 above explain your answer.**