INFORMATION RETRIEVAL – SHORT EXERCISES I – BOOLEAN RETRIEVAL AND NAVIGATIONAL PATTERNS

I. Consider the following documents **D1-D4** using 8 different terms:

D1 = {breakthrough drug schizophrenia}

D2 = {new schizophrenia drug}

D3 = {new approach treatment schizophrenia}

D4 = {new hope schizophrenia patient}

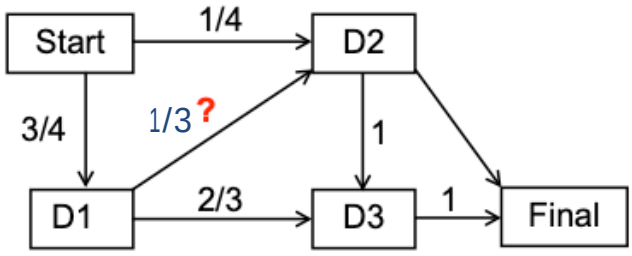
Fill in the term-document incidence matrix for this document collection.

	D1	D2	D3	D4
approach	0	0	1	0
breakthrough	1	0	0	0
drug	1	1	0	0
hope	0	0	0	1
new	0	1	1	1
patient	0	0	0	1
schizophrenia	1	1	1	1
treatment	0	0	1	0

What are the results returned for the below Boolean queries:

- schizophrenia AND drug — Answer: 1111 AND 1100 = 1100
- new AND NOT(drug OR approach) — Answer: 0111 AND NOT(1100 OR 0010) = 0111 AND NOT 1110 = 0111 AND 0001 = 0001

II. Given the following four sessions: {D1 D2 D3}, {D1 D3}, {D1 D3}, {D2 D3}, answer the questions related to using the Markov chain for mining navigational patterns.



What is P(D1→D2)? Answer: 1 - 2/3 = 1/3

What is the probability of P(Start→D1→D3)? Answer: 3/4 * 2/3 = 1/2

What is the probability of P(D3|D1)? Answer: 2/3 + 1/3 * 1 = 1