C. (6 points) Where are the relevant documents in the hit list? Mark a relevant document with an **R** in the corresponding box. Leave irrelevant documents unmarked.

1	2	3	4	5	6	7	8	9	10
R	**R**		**R**				**R**		**R**

11	12	13	14	15	16	17	18	19	20
							R		

Question 3. Boolean and Vector Space Retrieval (48 points)

Assume the following fragments comprise your document collection:

Doc 1: Interest in real estate speculation
Doc 2: Interest rates and rising home costs
Doc 3: Kids do not have an interest in banking
Doc 4: Lower interest rates, hotter real estate market
Doc 5: Feds' interest in raising interest rates rising

Assume the following are stopwords: an, and, do, in, not

A. (10 points) Construct the term-document matrix for the above documents that can be used in Boolean retrieval. The index terms have already been arranged for you alphabetically in the following table:

Term	**Doc 1**	**Doc 2**	**Doc 3**	**Doc 4**	**Doc 5**
banking	0	0	1	0	0
costs	0	1	0	0	0
estate	1	0	0	1	0
feds	0	0	0	0	1
have	0	0	1	0	0
home	0	1	0	0	0
hotter	0	0	0	1	0
interest	1	1	1	1	1
kids	0	0	1	0	0
lower	0	0	0	1	0
market	0	0	0	1	0
raising	0	0	0	0	1
rates	0	1	0	1	1
real	1	0	0	1	0
rising	0	1	0	0	1
speculation	1	0	0	0	0

B. (2 points each) What documents would be returned in response to the following queries?

interest NOT rates

Docs 1 and 3

(interest AND rates) NOT (rising OR kids)

(interest AND rates) → Docs 2, 4, 5
(rising OR kids) → Docs 2, 3, 5
(interest AND rates) NOT (rising OR kids) → Doc 4

((real AND estate) OR home) AND (interest AND rates)

((real AND estate) OR home) → Docs 1, 2, 4
(interest AND rates) → Docs 2, 4, 5
((real AND estate) OR home) AND (interest AND rates) → Docs 2, 4

(kids AND home)

None

Doc 1: Interest in real estate speculation
Doc 2: Interest rates and rising home costs
Doc 3: Kids do not have an interest in banking
Doc 4: Lower interest rates, hotter real estate market
Doc 5: Feds' interest in raising interest rates rising

stopwords: an, and, do, in, not

C. (20 points) Construct the vector space term-document matrix for the above documents (repeated from before) using *tf.idf* term weighting. Normalize your vectors. The following blank tables are provided for your convenience. You can use as many or as few of them as you wish. Clearly indicate your final answer.

		TF				
Term	**IDF**	**Doc 1**	**Doc 2**	**Doc 3**	**Doc 4**	**Doc 5**
banking	.699			1		
costs	.699		1			
estate	.398	1			1	
feds	.699					1
have	.699			1		
home	.699		1			
hotter	.699				1	
interest	0	1	1	1	1	2
kids	.699			1		
lower	.699				1	
market	.699				1	
raising	.699					1
rates	.222		1		1	1
real	.398	1			1	
rising	.398		1			1
speculation	.699	1				

TF.IDF

Term	Doc 1	Doc 2	Doc 3	Doc 4	Doc 5
banking			.699		
costs		.699			
estate	.398			.398	
feds					.699
have			.699		
home		.699			
hotter				.699	
interest					
kids			.699		
lower				.699	
market				.699	
raising					.699
rates		.222		.222	.222
real	.398			.398	
rising		.398			.398
speculation	.699				
length	.897	1.09	1.21	1.35	1.09

Normalized TF.IDF

Term	Doc 1	Doc 2	Doc 3	Doc 4	Doc 5
banking			.578		
costs		.641			
estate	.444			.295	
feds					.641
have			.578		
home		.641			
hotter				.518	
interest					
kids			.578		
lower				.518	
market				.518	
raising					.641
rates		.204		.164	.204
real	.444			.295	
rising		.365			.365
speculation	.779				

D. (4 points each) Simulate the retrieval of documents in response to the following queries. Indicate the order in which documents will be retrieved, and the similarity score between the query and each document.

interest rising

Doc 2: .365
Doc 5: .365
Doc 1: 0
Doc 3: 0
Doc 4: 0

real estate interest

Doc 1: .888
Doc 4: .59
Doc 2: 0
Doc 3: 0
Doc 5: 0

E. (2 points) Consider Doc 5: “Feds’ interest in raising interest rates rising.” Do the two instances of the term “interest” have the same meaning? What problem is this an example of?

Polysemy.