Natural Language Processing
Assignment 11
Type of Question: MCQ

Number of Questions: 10 **Total Marks: 10×1=10**

Question 1: One of your friends have recommended you to read the book "The God of Small Things". After reading the book, you want to summarize it. What kind of summarization method would you use for this purpose?

1. Abstractive single document summarization
2. Abstractive multi document summarization
3. Extractive single document summarization
4. Extractive multi document summarization

a. 1, 2
b. 3, 4
c. 1, 3
d. 2, 4

Answer: c
Solution:

Question 2: What kind of summarization approach is lexrank?

a. Extractive multi document generic
b. Extractive multi document query specific
c. Abstactive multi document query specific
d. Abstractive multi document generic

Answer: a
Solution:

Question 3: Identify whether the following statements are True or False.

1. Maximum Marginal Relevance strives to reduce redundancy while maintaining query relevance.
2. Query-focused summarization can be thought of as a complex question-answering system.

a. True, False
b. True, True
c. False, True
d. False, False

Answer: b
Solution:

For question 4-8, use the data given in Table 1.
Suppose you have trained a weather classifier with 5 classes - sunny, windy, rainy, foggy, and clear. Consider the confusion matrix shown in Table 1.

		Gold Labels				
		sunny	windy	rainy	foggy	clear
Predicted Labels	sunny	130	17	9	7	40
	windy	15	150	25	10	7
	rainy	10	45	150	23	5
	foggy	15	15	20	120	30
	clear	40	30	20	10	155

Table 1

Question 4: What is the macro averaged precision?

a. 0.6696
b. 0.6078
c. 0.6433
d. None of the above

Answer: c
Solution:

Question 5: What is the macro averaged recall?

a. 0.6464
b. 0.6540
c. 0.6190
d. None of the above

Answer: a
Solution:

Question 6: What is the accuracy of your classifier?

a. 0.6421
b. 0.6536
c. 0.6319
d. None of the above

Answer: a
Solution:

Question 7: What is the micro averaged precision?

a. 0.6915
b. 0.6421
c. 0.6245
d. None of the above

Answer: b
Solution:

Question 8: What is the micro averaged recall?

a. 0.6190
b. 0.6535
c. 0.6421
d. None of the above

Answer: c
Solution: Separate confusion matrix for each class is as follows:

class	TP	FP
	FN	TN

sunny	130	73
	80	815

windy	150	57
	107	784

rainy	150	83
	74	791

foggy	120	80
	50	848

clear	155	100
	82	761

recall $=\frac{TP}{TP+FN}$
precision $=\frac{TP}{TP+FP}$
accuracy $=\frac{TP+TN}{TP+FP+TN+FN}=\frac{705}{1098}\approx 0.6421$

macro averaged precision $=(0.6404+0.7246+0.6438+0.6+0.6078)/5\approx 0.6433$

macro averaged recall $=(0.6190+0.5837+0.6696+0.7059+0.6540)/5\approx 0.6464$

For micro averaged results, create pooled confusion matrix from all the classes.
micro averaged precision = micro averaged recall $=\frac{705}{705+393}\approx 0.6421$

Question 9: It is estimated that 40% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 98% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 3%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

a. 0.052
b. 0.048
c. 0.044
d. None of the above

Answer: c
Solution:

Let, A = Event that an email is detected as spam
B = Event that an email is spam
$P(B) = 0.4$
$P(B') = 0.6$
$P(A|B) = 0.98$
$P(A|B') = 0.03$

$$\begin{aligned} P(B'|A) &= \frac{P(A|B')P(B')}{P(A)} \\ &= \frac{P(A|B')P(B')}{P(A|B)P(B) + P(A|B')P(B')} \\ &= \frac{0.03 \times 0.6}{0.98 \times 0.4 + 0.03 \times 0.6} \\ &= \frac{18}{410} \\ &\approx 0.044 \end{aligned}$$

Question 10: Consider the system generated summary (S) and the reference summaries as follows:
S : the quick brown fox jumped over the lazy dog
R1 : the fox jumped over the dog
R2 : the brown fox leaped over the dog
R3 : the quick fox jumped over the lazy cat
What are the ROUGE-1 and ROUGE-2 recall values for the give summary with respect to the references?

a. 0.667, 0.500
b. 0.704, 0.417
c. 0.593, 0.333
d. None of the above

Answer: b
Solution:

ROUGE-1 recall = $\frac{6+6+7}{9+9+9} \approx 0.7037$
ROUGE-2 recall = $\frac{3+2+5}{8+8+8} \approx 0.4167$
