


National University of Computer and Emerging Sciences, Lahore Campus

	Course:	Information Retrieval	Course Code:	CS317
	Program:	BS(Computer Science)	Semester:	Fall 2019
	Duration:	25 Minutes	Total Marks:	12
	Paper Date:	22-Oct-19	Weight	3.3%
	Section:	B	Page(s):	2
	Exam:	Quiz 2	Roll No:	

Question 1:

How is the average precision related to precision recall curve? Show the relation using mathematical derivation [4 Marks]

Solution:

Average Precision approximates the area under precision recall curve.

Area under precision recall curve can be expressed using following integral.

$$\int_0^1 P(r) dr$$

Where $P(r)$ is Precision at recall r

Following summation is discrete approximation of this integral

$$\sum_{r=\{\frac{1}{R}, \dots, 1\}} P(r) \Delta r$$

Since Δr is constant so we can write it outside the summation

$$\Delta r \sum_{r=\{\frac{1}{R}, \dots, 1\}} P(r)$$

$\Delta r = 1/R$ where R = total relevant documents

$$1/R \sum_{r=\{\frac{1}{R}, \dots, 1\}} P(r)$$

This is formula of average precision

Question 2:

Compute mean average precision of following lists of documents for 2 queries. Leftmost is top ranked document. [4 Marks]

Q1 = R NR R NR NR R NR NR

Q2 = NR R NR NR NR NR R NR

Total relevant for query 1 = 8

Total relevant for query 2 = 15

Solution:

Average Precision for Q1 = $(1 + 2/3 + 3/6)/8 = 0.27$

Average Precision for Q1 = $(1/2 + 2/7)/15 = 0.05$

Mean Average Precision = 0.32

Question2:

Consider following collection of 3 documents. [4 Marks]

Document	Words
D1	a b b a b b c
D2	a a b a b a
D3	b b b b b b c c

Query = < a b >

Use Witten Bell smoothing to find similarity of document D3 with query.

Solution:

For document D3

$N = 8, V = 2$

$\text{Prob}(a) = N/(N+V) * 0 + V/(N+V) * 6/21 = 0 + 2/10 * 6/21 = 0.057$

$\text{Prob}(b) = N/(N+V) * 6/8 + V/(N+V) * 12/21 = 8/10 * 6/8 + 2/10 * 12/21 = 0.71$

$\text{Prob}(a,b) = \text{Prob}(a) * \text{Prob}(b) = 0.04$