

V3

National University of Computer & Emerging Sciences

FAST-Karachi Campus

CS4051- Information Retrieval

Quiz#2

Dated: March 13, 2023

Marks: 20

Time: 20 min.

Std-ID: _____ Sol _____

Question No. 1

Consider the output of an IR system against a query q. It returns the following rank order of the documents: Result-Set = d4, d6, d3, d2, d7, d1, d5, d9

Following are the relevance scores of all the relevant documents in the corpus. The score is 5 to 1, and 5 being the most relevant and 1 means – non-relevant.

Relevance Scores: d1(3), d2(3), d3(4), d4(5), d5(5), d6(4), d7(2), d8(2), d9(1), d10(3).

Answer the following questions: [2.5 X 4]

- a. What is the Cumulative Gain at 5?

Rank order – $R_5 \{d4, d6, d3, d2, d7\}$

$$CG_5 = Rel(d4) + Rel(d6) + Rel(d3) + Rel(d2) + Rel(d7) = CG_5 = 5+4+4+3+2 = 18$$

- b. What is the Discount Cumulative Gain at 5?

Rank	Rel(d_i)	Position Gain [$\log(i+1)$]	Score = Rel(d_i)/Position Gain
1 -> d4	5	1	5
2 -> d6	4	1.585	2.52
3 -> d3	4	2	2
4 -> d2	3	2.322	1.29
5 -> d7	2	2.585	0.77

$$DCG \text{ at } 5 = 5+2.52+2+1.29+0.77 = 11.58$$

- c. What is the idealized Discount Cumulative Gain (IDCG) at 5?

In this scenario Idealized Discount Cumulative Gain is equal to DCG at 5

- d. What is the normalized Discount Cumulative Gain (nDCG) at 5?

Normalized Discount Cumulative Gain (nDCG) at 5 = 1 as Both DCG and IDCG are equal.

Question No. 2

What is a Test Collection in the context of Information Retrieval? What are its essential components? [5]

To measure ad hoc information retrieval effectiveness in the standard way, we need a test collection consisting of three things:

1. A document collection
2. A test suite of information needs, expressible as queries
3. A set of relevance judgments, standardly a binary assessment of either relevant or non-relevant for each query-document pair.

Question No. 3

Assume there are 16 total relevant documents in a collection. Consider the following result against a query that return 10 documents in the given order., where plus indicates relevant and minus indicates non-relevant: {+, +, -, -, -, -, -, +, +, +}

Calculate the following evaluation measures on the ranked list.? [2.5 x 4]

a. What is the Precision?

$$\text{Precision} = \# \text{ relevant retrieved} / \text{total retrieved} = 5/10 = 0.5$$

b. What is the Recall?

$$\text{Recall} = \# \text{ relevant retrieved} / \text{total relevant} = 5/16 = 0.312$$

c. What is the F1 Score?

$$F1 = (2 * P * R) / (P + R) = (2 * 0.5 * 0.312) / (0.5 + 0.312) = 0.312 / 0.812 = 0.384$$

d. What is Average Precision?

$$AP = (1 + 1 + 3/8 + 4/9 + 5/10) / 5 = 3.319 / 5 = 0.66$$