## 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: | A | Page(s): | 2 |
| Exam: | Quiz 2 | Roll No: | |

**Question 1:**

**TRUE or FALSE. Justify your answer. [1 Mark]**

1. Mean average precision is biased by queries with more relevant documents. **[2 Mark]**

**TRUE or FALSE**
**Justification:**
FALSE
While computing MAP we divide the accumulated precisions with total number of relevant documents. The MAP of queries with large number of relevant documents will be penalized by division with a large number.

2. Assume we use Dirichlet Prior smoothing; duplicate the document content multiple times will not change the resulting smoothed document language model. **[2 Mark]**

**TRUE or FALSE**
**Justification:**
FALSE
It will change the smoothing weight because the smoothing weight is a function of document length and document length will change.

3. Given a well-tuned unigram language model $p(w|\theta)$ estimated based on all the text books about the topic of "information retrieval", we can safely conclude that
$p(\text{"information retrieval"}|\theta) > p(\text{"retrieval information"}|\theta)$. **[1 Mark]**
**TRUE or FALSE**
**Justification:**
FALSE

Unigram model does not measure the word ordering so the probability of "information retrieval" and "retrieval information" will be same.

**Question 2:**

Compute NDCG @ 5 of following list of documents. Leftmost is top ranked document. [4 Marks]
1, 3, 2, 0, 2, 1, 1, 0 , 3, 2, 1
All relevant documents in collection =  {3, 3, 2, 2, 2, 2, 1, 1}
**Solution:**

 Discounted Gain  = 1, 3/lg 2, 2/lg 3 , 0 / lg 4, 2 / lg 5,
Discounted Gain  = 1, 3, 1.26, 0, 0.86

Discounted Cumulative Gain = DCG = **1, 4, 5.26, 5.26, 6.12**

Compute DCG of ideal list {3, 3, 2, 2, 2, 2, 1, 1} using same method and then divide the DCG with ideal DCG

DG of Ideal = 3, 3, 2/lg 3, 2/ lg 4, 2/ lg 5  = 3, 3, 1.26, 1, 0.86

DCG of Ideal list = **3, 6, 7.26, 8.26, 9.12**

NDCG = 1/3, 4/6, 5.26/7.26, 5.26/8.26, 6.12/9.12  = 0.33, 0.67, 0.72, 0.64, 0.67

**NDCG @ 5 = 0.67**

**Question2:**

DCG scores of two different queries for a search engine are as follows:
DCG score for Query 1 = 4.5
DCG score for Query 2 = 2.5
Can we infer that search engine's performance for query 1 is better than query 2. Justify your answer. [3 Marks]

**Solution:**

No, search engine's performance for query 1 is not necessarily better than query 2 because this is DCG score and we have not considered the total number of relevant documents for both queries. It is possible that query 1 has a lot relevant documents with high relevance grades as compared to query 2. NDCG is a better measure for making fair comparison between different queries.