## **CSCI 6370 IR and Web Search**

## Homework Assignment Two

Total points: 100. Due: Monday, June 15.

### Problem 1. [50 points]

This assignment is designed for you to get familiar with various evaluation mechanisms in IR. Given the set of data in the table,

1. Compute the recall/precision rate at each relevant document point.
2. Compute F-Measure at each relevant document point.
3. Compute the corresponding E-measure with β = 0.8 at each relevant document point.
4. Compute the R-Precision.
5. Computer the average precision, which is the average of the precision values at the points at which each relevant document is retrieved.

The total relevant documents in a document set is 5. The top 10 retrieved documents are

listed with the ones that are relevant marked.

|  |  |  |
| --- | --- | --- |
| Order | Doc # | Relevant |
| 1 | 586 | X |
| 2 | 357 |  |
| 3 | 358 | X |
| 4 | 108 | X |
| 5 | 345 |  |
| 6 | 114 |  |
| 7 | 555 | X |
| 8 | 888 |  |
| 9 | 860 |  |
| 10 | 167 | X |

### Problem 2. [25 points]

Suppose we use the following the method to reformulate the query vector in response to relevance feedback:

Consider the initial query vector is

The relevant feedback gives two relevant vectors d\_1 and d\_2 and one irrelevant vector d\_3 as follows:

Calculate the reformulated query vector .

### Problem 3. [25 points]

Phrasal query is to retrieve documents with a specific phrase (ordered list of contiguous words). For example, phrasal query for Q=”computer learning theory” needs to retrieve all documents containing the phrase “computer learning theory”. We usually ignore cases of the letters in the query. We may also allow intervening stop words and/or stemming. Describe an algorithm to do phrasal query with help of an inverted index.