Problem 1 [12 points]. Compute the similarity between query \( q_3, k_1 \land (k_2 \lor \neg k_3) \), and document (doc) \( c \) using (i) the Fuzzy Set IR model, where doc \( C \) is one of the 5 documents (i.e., \( A, B, C, D, \) and \( E \)) and (ii) the occurrences of the following set of keywords:

<table>
<thead>
<tr>
<th>Documents/Index Terms</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k_1 )</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>( k_2 )</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>( k_3 )</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

where \( Y \) (\( N \), respectively) in the table denotes that the respective document contains (does not contain, respectively) the corresponding index term.

Problem 2. An IR system returns eight relevant documents and ten non-relevant documents. There are a total of twenty relevant documents in the collection.

(a) [6 Points]. What is the precision of the system on this search, and what are its recall and harmonic mean (i.e., F-measure)?

(b) [2 Points]. What is the advantage of using the harmonic mean rather than “averaging” (using the arithmetic mean)?

Problem 3. Given below is a table showing how two human judges rated the relevance of a set of twelve documents retrieved by an IR system in response to a user’s query \( Q \) in which ‘1’ stands for relevant and ‘0’ for non-relevant. Further assume that there are eight relevant documents with respect to \( Q \) in the collection. (Note: documents that are relevant are indicated in the following table with their Doc_ID bold.)

<table>
<thead>
<tr>
<th>Doc_ID</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Judge 2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

(a) [8 Points]. Draw the Interpolated Recall-Precision Graph based on the rating given by each of the two judges, respectively.

(b) [2 Points]. What are the false positives and false negatives based on the rating of each of the judges, respectively?

Problem 4. Consider an information need for which there are four relevant documents in the collection for the two queries \( q_1 \) and \( q_2 \). The top ten results retrieved by the system for each query are shown below where \( R \) stands for relevant and \( N \) for non-relevant, and the leftmost item is the top ranked search result.

Query 1 (\( q_1 \)): \( R \ R \ N \ N \ R \ N \ N \ N \ R \)
Query 2 (\( q_2 \)): \( N \ R \ N \ R \ N \ R \ N \ N \ R \ N \)

---

Homework Assignment 6
due Monday, June 6
(a) [3 Points]. What is the MAP of the system?

(b) [1 Point]. Why is it important in getting a good MAP score?

(c) [3 Points]. What is the Mean Reciprocal Rank (MRR) of the system?

(d) [3 Points]. What is the Precision at 5 of $q_1$ and $q_2$, respectively?

Problem 5. Assume that four relevant documents are retrieved for each one of the two queries $q_1$ and $q_2$, and the sets of top ten results retrieved for $q_1$ and $q_2$ are given as follows (with the leftmost item being the top ranked search result):

\[
q_1: \text{R} \; \text{N} \; \text{R} \; \text{N} \; \text{N} \; \text{N} \; \text{N} \; \text{R} \; \text{R} \; \\
q_2: \text{N} \; \text{R} \; \text{N} \; \text{N} \; \text{R} \; \text{R} \; \text{R} \; \text{N} \; \text{N} \; \text{N}
\]

where $R$ stands for relevant (document) and $N$ denotes non-relevant (document). Further assume that binary relevance scale on the results is used.

(a) [5 points]. What is the DCG at 5 and 10 for $q_1$ and $q_2$, respectively?

(b) [5 points]. What is the NDCG at 5 and 10?