1 Search engine evaluation

1. Below is a table showing how three human judges rated the relevance of a set of 12 documents to a particular information need (0 = nonrelevant, 1 = relevant). Let us assume that you’ve written an IR system that for this query returns the set of documents \{4, 5, 6, 7, 9\}.

   a) Calculate the kappa measure between the three judges.
   b) Calculate precision, recall, and $F_1$ of your system if a document is considered relevant only if the three judges agree.
   c) Calculate precision, recall, and $F_1$ of your system if a document is considered relevant if either judge thinks it is relevant.

   \[
   \begin{array}{ccc}
   \text{DocID} & \text{Judge 1} & \text{Judge 2} & \text{Judge 3} \\
   1 & 0 & 1 & 0 \\
   2 & 0 & 1 & 1 \\
   3 & 0 & 1 & 1 \\
   4 & 0 & 1 & 1 \\
   5 & 1 & 0 & 1 \\
   6 & 1 & 0 & 1 \\
   7 & 1 & 0 & 0 \\
   8 & 1 & 0 & 0 \\
   9 & 1 & 1 & 1 \\
   10 & 1 & 1 & 0 \\
   11 & 0 & 0 & 0 \\
   12 & 0 & 0 & 1 \\
   \end{array}
   \]

2. A collection of documents contains 20 documents that are relevant for a given query. For this query, the search engine returns 8 relevant and 10 not relevant documents.

   a) Compute precision and recall of this search result as well as the $F_1$-measure.
   b) What are the advantages of the $F_1$-measure compared to the arithmetic mean of precision and recall?

3. A collection of documents contains 10 documents that are relevant for the query $q$. For this query, the search engines $S_1$ and $S_2$ return the following relevant (R) and non-relevant (N) documents:

   $S_1$: NNNRR NRNRR

Sergej Wildemann - wildemann@L3S.de
$S_2$: RRNNN  NNRRN

Draw a precision-recall diagram for the both search results and compare the quality of the search results based on the **interpolated precision at 25% recall**.

<table>
<thead>
<tr>
<th>Recall(%)</th>
<th>$P(S_1)$</th>
<th>$P(S_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>