<table>
<thead>
<tr>
<th>Domain Task assessment</th>
<th>Domain Task assessment</th>
<th>Domain Task assessment</th>
<th>Domain Task Assessment</th>
<th>Domain Task assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which categories of plagiarism appear in the document?</td>
<td>Which of the categories of plagiarism appear most frequently?</td>
<td>What are the removed/inserted/modified words within the suspicious text segment in comparison to the original source?</td>
<td>(specific text segment) For a given pair (source/destination) how is the text different? What has been added? removed?</td>
<td>(summary of differences) For a given pair (source/destination) how is the text different? What has been added? removed?</td>
</tr>
<tr>
<td>Domain Task verification</td>
<td>Domain Task verification</td>
<td>Domain Task verification</td>
<td>Domain Task verification</td>
<td>Domain Task verification</td>
</tr>
<tr>
<td>What is the evidence of plagiarism? (stats to assess plagiarism decision &amp; decision source -- human, algorithm, etc.)</td>
<td>How much overlap is there between source and destination (specific instance)? (characters versus percent)</td>
<td>What is the distributional overlap (size/amount/percent) between all pairs of suspected copies?</td>
<td>What is the central tendency (mean, mode, median) of overlap size/amount/percent?</td>
<td>What are the structural differences and similarities between the suspicious text and the original source?</td>
</tr>
</tbody>
</table>
What are the textural differences and similarities between the suspicious text and original source?

Which are the text segments that are suspicious for plagiarism?

How are the suspicious text segments distributed among the pages of the entire document?

How long are the original/suspicious texts?

Where are the suspicious text segments?

Who is/are the author(s) of the original/suspicious documents?
How many sources are used?

How many sources are cited?

Which sources are used most for paraphrasing text?

What is the average length of the suspicious text segment?

What is the distribution of the length of suspicious text segment across the document?

How many suspicious text segments are detected?

How are different pieces of a source document plagiarised? (distribution from source to destination)