Assignment 1
Due on 2019-09-14, 23:59 IST.

Check all the right answers.

1. Describe pre-processing steps in an NLP application:
   - Tokenization
   - Substitution
   - Stemming
   - Part-of-speech tagging
   - Named entity recognition
   - Feature extraction
   - n-gram extraction
   - Sentence splitting
   - Dependency parsing

2. Pick the remaining sections:
   - Case, or, and, is
   - Help, in, is, the
   - Instance, looking, inside, study
   - Include, looking, inside, study
   - Include, looking, inside, study
   - All of the above
   - No, the answer is incorrect.
   - No, the answer is correct.
   - The answer is incorrect.
   - The answer is correct.

3. Clear tracking is used for:
   - Hyperlink tracking
   - Sentence tracking
   - Dependency tracking
   - Lexicon tracking
   - All of the above
   - No, the answer is incorrect.
   - No, the answer is correct.
   - The answer is incorrect.
   - The answer is correct.

4. Consider a corpus with 3000 documents. The word token matrix in some documents has the following frequency:

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>100</td>
</tr>
<tr>
<td>b</td>
<td>200</td>
</tr>
<tr>
<td>c</td>
<td>300</td>
</tr>
<tr>
<td>d</td>
<td>400</td>
</tr>
<tr>
<td>e</td>
<td>500</td>
</tr>
</tbody>
</table>

   What is the number of unique words in the document?
   - 400
   - 500
   - 50
   - 100
   - 3000
   - No, the answer is incorrect.
   - No, the answer is correct.
   - The answer is incorrect.
   - The answer is correct.

5. Consider the following corpus of 4 documents:

   Document 1: Scape
   Document 2: Scape
   Document 3: Scape
   Document 4: Scape

   Calculate the Tfidf for the word 'scape' in the given documents:

   1. \( \text{TF-IDF} = \frac{\text{Term Frequency}}{\text{Inverse Document Frequency}} \)

   - No, the answer is incorrect.
   - No, the answer is correct.
   - The answer is incorrect.
   - The answer is correct.

6. A college student wants to be less conclusive in his term paper, but keeps ending out of a word, which ends up being out of a word. He tries to fix the grammar with an "r". What is the correct way to improve his grammar?
   - Use a spell checker
   - Use a grammar checker
   - Rewrite the sentence
   - Use a thesaurus
   - No, the answer is incorrect.
   - No, the answer is correct.
   - The answer is incorrect.
   - The answer is correct.