

Instructions for NLP Practical (Units of Assessment)

Experiments concerning Text Understanding of IELTS test questions

Simone Teufel (Lead demonstrator Guy Aglionby)
sht25@cl.cam.ac.uk; ga384@cl.cam.ac.uk

This is the third part of the NLP practical, which is concerned with text understanding. In particular, the scenario is to simulate a second language learner's performance on the IELTS test.

You are given two texts, Text 1 and Text 2, and the corresponding questions. (Question 5 of Text 1 is made up by me, but all others are real IELTS training material). Please read them and answer them for yourself.

In the following, you will be stepped through 4 small (thought) experiments, which you can answer with minimal experimentation, entirely without programming if you like, and with some thinking. Please answer each task in 250 words. For Task 4 you may exceed this word limit if you have any great ideas, but a reasonable design can be described in 250 words.

- **Task 1** tests your knowledge of syntactic dependencies. Consider the following question to Text 1: "What sort of water are you advised to use?" Show which dependencies from the parsed version can be used to restrict the possible answers to the correct one. Now do the same for Question 12 of Text 2.
- **Task 2** will encourage you to think about two different metrics for lexical similarity – semantic space-based similarity and ontology-based similarity. "What should you do if your iron starts to drip water?" The word "drip" does not occur in text, so show how you would use Wordnet and, separately, google-supplied W2V embeddings to find the most similar word to it in text. Again, demonstrate with another example of your choice.
- **Task 3** is about chains of reasoning, i.e., abstract connections between words and non-lexical knowledge that leads to the solution in small steps.

For instance, the statements

```
‘‘clothes worn by a person are in direct touch with skin’’
```

and

```
‘‘skin can burn if touched by a hot surface’’
```

are examples of statements that might form part of the reasoning chain that answers Question 5 of Text 1: "Which misuse of the iron could result in a person getting hurt?". Please spell out this reasoning chain. Is this a task that can be implemented? If yes, state how. If no, state the largest obstacle.

- **Task 4** You are now given a near-impossible task: the task of automating the answering of IELTS questions. Nobody expects it to be perfect, but you are asked to provide "a workable compromise that will not embarrass the company too much", and let's say, a resource of 2 person years to put this plan into action. Which existing NLP technology can you rely on? Where would you spend most of the programming time?

Please submit both Reports 2 and 3 by the deadline stated on the website.