

Instructions for NLP Practical (Units of Assessment)

Assignment 3

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. In line with the Tokyo University/NII project, an important specification aspect of the system is that its internal workings should provide explainability for any result.

You are given two texts, Text 1 and Text 2, and the corresponding questions in pdf/png and parsed versions. (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. (Which steps did you perform to solve the tasks?)

In the following, you will be stepped through 3 small (thought) experiments, which you can answer by thinking and with minimal experimentation, entirely without programming if you like. Please answer each task within a 333 word limit.

- **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** encourages you to think about different metrics for lexical similarity – e.g. semantic space-based similarity and ontology-based similarity.

Treat Question 3 from Text 1 and one other question of your choice. For instance, in Question 3 (“What should you do if your iron starts to drip water?”), the word “drip” does not occur in text. Show how you would use Wordnet and, separately, google-supplied W2V embeddings to find the most similar word to it in text. Choose your other question you treat in such a way that you can best demonstrate your knowledge of semantic similarity metrics.

- **Task 3** is concerned with 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.

Then argue whether you think the task can be simulated automatically in a robust way, and why. If your solution involves ML (it doesn’t have to), then consider that the availability and adequacy of existing training material (do you have to create more?), and how the entire system could provide the explainability required in the specification for the system. You may consider what the biggest obstacles are, and how they could be overcome. Give evidence for your answers.

Important: In all three tasks, please demonstrate knowledge of NLP by **citing relevant Overview of NLP lectures**.

Please submit both Report 3 by the deadline stated on the website.