

NLP Interview Task

December 2022

1 Overview

For the research position interview we are asking that candidates prepare two items:

1. Two slides introducing themselves, their research history, and interests.
2. Two slides describing their approach and findings for a small NLP task (described in this document).

2 NLP Task

2.1 Dataset

Download the following datasets:

1. https://raw.githubusercontent.com/reality/intask/main/pt_notes.csv
2. <https://raw.githubusercontent.com/reality/intask/main/patients.csv>

They describe a set of textual clinical notes describing patient admissions in a critical care setting.

2.2 Task

Perform an exploratory stratification of biomedical entities based on the text, such as admissions, diseases, or phenotypes, with a focus on exploring hypothetical relationships that can be discovered from the text. The presentation should focus on methodology, findings, and next steps.

2.3 Conditions

- Please do not spend more than 2 hours on the task, and do not provide more than 2 slides describing your approach and results. The task is intended as a jumping-off point for discussion during the interview, and as such more time spent on the problem will not confer a greater advantage.
- Any software environment or toolset you want to use is okay.