# SMU MSDS 7337 Mid-Term Exam Summer 2019

Korey MacVittie

# Short Essay Responses

(25 pts each, 300-500 words each).

1. Select one career or industry that makes use of applied NLP.
   1. Explain generally how that field or career utilizes NLP.
   2. Explain at least some methods of NLP that are very likely to be used in the career or industry you selected.
   3. Give at least one specific example of a use case for NLP within the chosen field, and explain how the problem or situation is (or could be) improved by applying NLP.

Machine translation is a field that makes significant use of NLP, by translating words or sentences in one language to another. While single word translation can be relatively straight-forward, sentence translation is significantly more complex. The machine must be able to understand a word's meaning, derived from context (for instance, the word "read" in English has multiple meanings, which must be derived from the sentence as a whole); it then must locate words that match or convey a similar thought in the target language, and reconstruct the sentence using the target language's grammar.

The primary means of doing this today involves tagging: identifying which part of speech a particular word is, to assist in both constructing the appropriate grammar in the target language, as well as determining which meaning of the word is to be used. This may also involve stemming: the process of finding the root word in a word (for instance, in “finding,” the root word is “find”), which is an important part of determining both the meaning and type of a word: rather than have many definitions for the word “find” and its variations, a dictionary could simply have “find” and list out its various possible endings, each indicating what type of speech it is, which would help retain the meaning of the word.

Translation as a whole benefits significantly from NLP. Even in its current state, we often find ourselves reliant upon individuals with the proper linguistic knowledge: for instance, I sometimes make use of English-French machine translation, but find myself making small adjustments to the result using my minimal French knowledge, and wouldn’t rely upon it for professional use. In a world in which it is far easier for conversations to happen between individuals who do not share a language, having tools that allow us to communicate readily will become more and more important, given the difficulty of learning new languages.

1. Choose one of the “trade-offs” in NLP that was covered in the materials for this course.
   1. Explain the trade-off in general terms. Define the two choices.
   2. Explain the benefits and weaknesses of each side of the trade-off. Include at least one benefit and one weakness of each.
   3. Describe a work-situation that would make one of the choices in the trade-off much better, in terms of practical outcomes for you and your stakeholders on a project.

# Essay Response

(50 pts, 600-1000 word essay).

1. Describe the overall NLP “pipeline” for a text analytics project as described in the materials of this course. First give an overview, then briefly describe and explain each phase of the pipeline.
2. For each phase of the pipeline, either (a) recommend a specific tool for getting the job done, e.g. a Python package, and a reason why you recommend it, or (b) explain a choice that must be made, where you would configure the process for that step very differently depending on the kind of application.

Note: For 2, do each of (a) and (b) at least once.