Team Name: Muskrat Attack
Team Member Name: Tomas Anthony
NetID: tomasaa2
Team Captain: Tomas Anthony
Chosen System: System Extension
Sub Topic: ExpertSearch - *Extracting relevant information from faculty bios*

Datasets:
I plan on using mined text from faculty bios, previously performed in MP 2.

Algorithms and Techniques:
I plan on using BERT based NLP techniques to fine tune a NER classifier model to better structure and retrieve information from the faculty bios.

I will demonstrate using F1-Score and other metrics for the classifier models performance that the NER exceeds the performance of the Stanford boilerplate NER.

My code will communicate with the system by being integrated into the code base in the same way the current NER system is. If necessary, due to Github size limitations, I will upload the ML NLP model to a hosting service to be interacted with from there.

Programming Language:
Python

Workload Justification:
I am a single person and will need at least 20 hours to data mine and fine tune a classifier model. Data mining could take between 5-10 hours. Coding the classification model could take another 10 hours. Integrating this into the existing ExpertSearch code base could take another 5. Logistic work and miscellaneous coding could be 5-10 hours.