

HW3 Report

Huilan You

Task 1:

The goal of this task is to find mismatches between the number of authors and email addresses. In order to solve this task, I used the functions (`load_map`, `get_entry_dict`) we learn in class to load the bib files, format them into data dictionary in Python. Then, I changed the function `get_email_dict` to extract missing emails for different patterns. I found three more patterns that we can extract the emails. For different patterns, I split them differently and combined them back and put the email to a list. The final results called `you_mismatch.tsv`.

Task 2

The goal of this task is to allocate weights to the institutions based on the school of the authors. I created two functions to accomplish the goal. The first is called `get_domains`. As the domains of the email often represent different institutions, I collect domains name first. Second, based on the domains, I used `get_weights` to calculate the weights for different institutions.