Kristen McGarry

March 8, 2017

LHS 712 Progress Report

***Recap on Project Goals***

For my NLP semester long project, I wanted to explore help forums from the MedHelp site that specifically related to dementia and Alzheimer’s. Through these forums, I’m hoping to identify the needs of the memory loss community, but also to be able to distinguish between the Alzheimer’s community and the dementia community, given that these terminologies are commonly used interchangeably. To help organize my thoughts on the progression of my project, I have included an overview diagram in figure 1(see page 2).

***Current Progress & Next Steps***

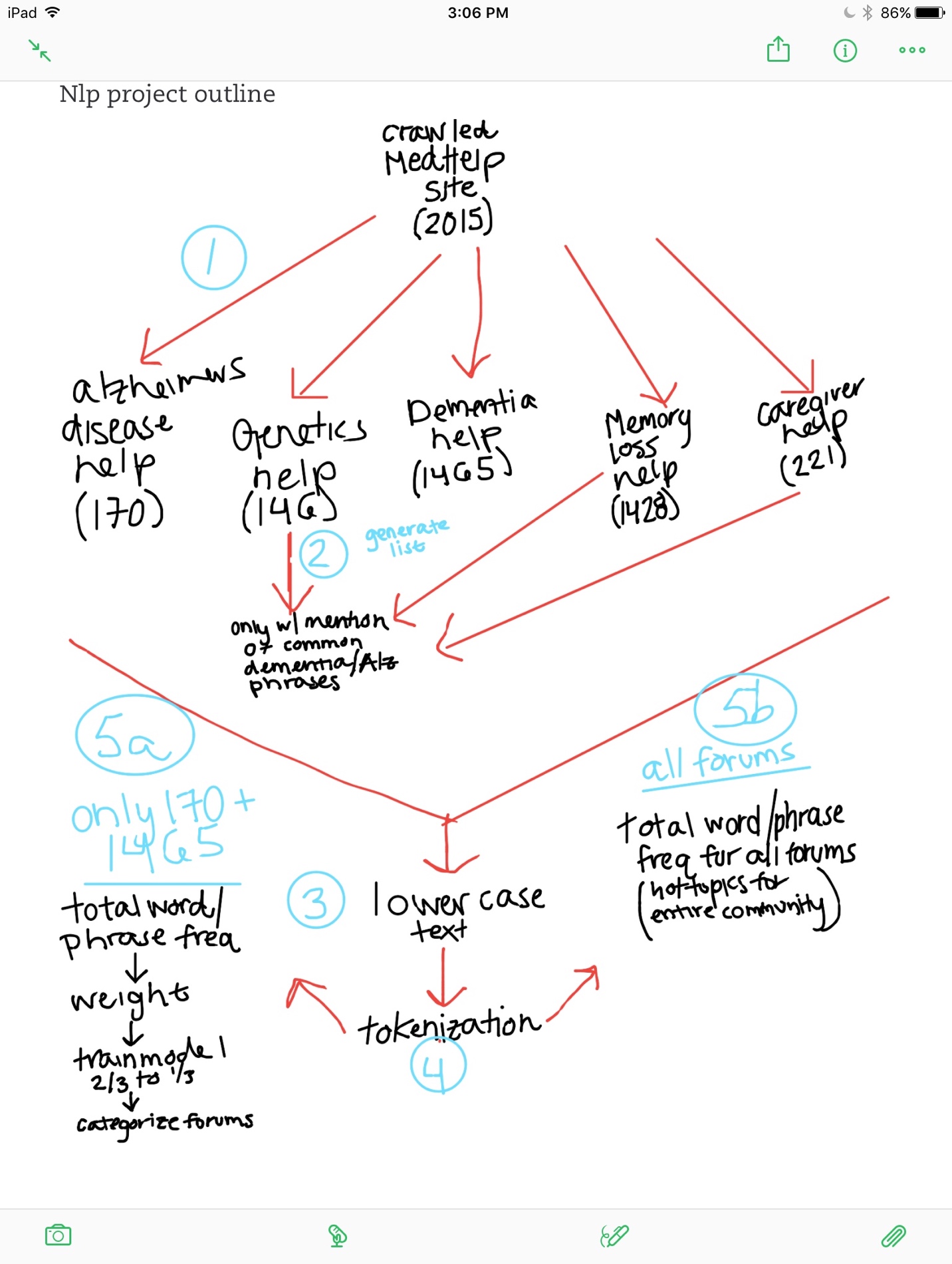
In relation to the diagram below, I have completed step 1, taking the data collected from the previously crawled MedHelp site, and extracted information from the following forums into individual text files:

* Alzheimer’s Disease Help (forum id: 170)
* Genetics Help (146)
* Dementia Help (1465)
* Memory Loss Help (1428)
* Caregiver Help (221)

For now, I have decided to work with the questions data, because I’m not sure what additionally features the comments could provide at this point. As far as data cleaning (step 1 & step 2), Alzheimer’s Disease Help and Dementia Help provide specific questions regarding these communities so not much cleaning is needed there. I am planning on removing erroneous questions that seem misplaced (unsure yet how to do this not manually). As for the other forums (Genetics, memory loss, and caregiver), I need to extract the questions that are pertinent to the dementia and Alzheimer’s community (step 2). To do this, I’m in the process of identifying a list of key, identifying words that are commonly used, and if these words are present in the text, including that question.

After I’ve completely cleaned and organized the data, then I will proceed to steps 3-5. In step 3, I have proposed to lower case all text, since I’m looking for key words/phrases and not sentence structure, and in step 4 perform tokenization. Next, I differentiated the comparison of dementia and Alzheimer’s (including ability to differentiate questions) (step 5a) and the community, as a whole (step 5b).

At this point, I believe I’m on track to have the project be completed in mid-April. However, I would like to meet with you soon to walk through my thought process and seek input before I get too deep in the process (I’ll email you separately).



**Figure 1. Outline of NLP Project Progression.**