Project manager: Himanshu Agrawal
Team: Zhen Rong, Liew (Dennis), Steven Garcia, Mark Simon, Asa Soriano

Our project in general will be more exploratory in nature. We will mainly be focused on utilizing natural language processing techniques to distinguish between causal incidents in Verbotic’s medical code. Himanshu has advised us on using word vectors and machine learning to evaluate the semantical meaning of a given word sentence (in English). The chunk of the project will involve researching and/ or implementing online libraries to construct an efficient algorithm to achieve the main goal stated above.

We have subdivided the goal as follows:

1. Implement an algorithm where it receives two inputs (sentences of any length). Then determine the similarity in semantic meaning between the two. Generate an output between 0.0 and 1.0.

   For example, the sentence “Bob crashed his car into a tree” compared to “A man has smashed his auto mobile into a plant” should score higher in similarity than if compared to “Alice went swimming in the ocean”.

2. Iterate on the previous algorithm to accept any number of sentences. This can be in the form of any text file. The algorithm will need to go through all the sentences, accurately distinguish where a sentence starts and ends, and ultimately return the two most similar sentences found within the input text.

3. Apply algorithm (2) on Verbotic’s medical code – working primarily on the causal incidents.

   We will be using the medical guidelines codebook to further train the word vector model to better identify semantical meaning regarding medical terms. This will be discussed in more detail with Himanshu when we complete goal 2.