

Kim Soffen  
Lisa Ma  
Hsiu-Chi (Jimmy) Lin  
CS171  
April 17th 2015

### Final Project Design Studio: Peer Feedback

We would like to thank **Dylan Tan**, **Fanney Zhu**, and **Jack Wongtam** for their constructive feedback about our visualization sketches and offer our appreciation for their fair and honest opinions. Our project, from the start, has been firmly purposed to provide a useful, practical, and interesting analysis of word usage trends in the Harvard Student publication, *The Crimson*, so their interests as students in the data are particularly important to us.

We were happy that they liked the idea of looking at word usage frequencies in past publications to get a sense of the issues of concern for the Harvard Community over time. They were concerned that our plans for a multiple line graph may eventually cause "hairball" issues, with too many data lines obscuring useful and interesting information. We have since then adjusted by placing a hard cap (currently 5) on the number of words that can be displayed at any time and have implemented hover functionality to give clarity as to which line represents the frequency of each word.

From there, they raised concerns about how large our dataset would be. Since *The Crimson* is a well established publication, their articles reach back in significant quantity to the early 1900's and with articles interspersed between 1700 and 1900. The amount of data we acquired from data scraping *The Crimson* lead to a file 1.2 GB large. We decided to accommodate this concern by either preprocessing the data so that visualization was manageable and relatively swift or to limit the visualization to 500 of the top words (judged by frequency). We believed that our fix was reasonable, given that the visualization was to give an idea of overarching trends over time, so the most significant words would likely also be the most defining of their contemporary era.

Last, they said they were interested in looking at interactions between different words instead of simple frequencies. We will attempt to satisfy this desire by scraping further to find linked associations between different words and implementing a creative layout (possibly a linked node layout, as suited for network visualizations) to demonstrate correlations between word appearances over time.

Overall, we received praise for a practical design meant to clearly display data as a general trend over time but we are definitely considering various methods to show interaction and strength of interaction of words in a meaningful and effective way.