

6 January 2014

To the NSF Review Panel:

I am writing in support of the proposal titled CHS: Small: Interactive Machine Learning for Text Analysis, submitted by Jeffrey Heer.

For background, I joined the Department of Human Centered Design & Engineering at the University of Washington as an Associate Professor in 2011, after six years as a computational scientist at Lawrence Berkeley Lab. I earned my Ph.D. in computer science from UC Berkeley in 2004 with Professor Marti Hearst. My current research focuses on HCI, visual analytics, statistical analysis of affect in text communication, and the visual understanding of very large data sets. I have authored or co-authored over 60 peer-reviewed publications and over 100 other publications in multiple domains including HCI, visual analytics, data science, machine learning, and astrophysics. In 2008, I received the Presidential Early Career Award for Scientists and Engineers (PECASE) for my work in visual analytics for data-intensive science.

As someone who is familiar with the visualization research literature, I consider Heer's work consistently among the best in the field. Unlike many of the research studies that are done in data visualization, his work always addresses useful topics and produces practical results that solve real problems. Further, his publications are always models of clarity, often elucidating new insights with precision and intelligence, and functioning as trailblazers that frequently define the fundamentals of a domain.

I am excited that Prof. Heer is moving into the field of machine learning for text analysis, as I am certain he will make groundbreaking contributions in this area. I am eager to collaborate with him on this work, as my research group is currently developing machine learning and visual analytics tools to detect and analyze emotion in text-based online communication. We recently published a paper at CSCW 2013 (Brooks et al., "Statistical affect detection in collaborative chat") that outlines our approach. The detection of emotion in text is a process that has typically been studied through the manual labeling of lines of text. We developed a process to scale this type of qualitative research to large-scale social media data sets by building a pipeline of natural language processing and machine learning techniques to build automated classifiers of emotion in text. Taking a novel dynamic approach to emotion detection in chat, we achieved excellent results with our pipeline and released it as an open source tool. We are enthusiastic about collaborating with Prof. Heer's research group in this area, as I'm certain that his techniques will strengthen our own approach and our tools for emotion detection could potentially be incorporated into his more general text analysis pipeline.

Sincerely,

Cecilia Aragon

Cecilia R. Aragon

Associate Professor, Human Centered Design & Engineering
University of Washington