

Keynote

Automating Analysis of Social Media Communication: Insights from CMDA

Susan C. Herring

School of Library & Information Science and Department of Linguistics

Indiana University

Bloomington

`herring@indiana.edu`

Abstract

A growing body of research analyzes the linguistic and discourse properties of communication in online social media. Most of the analysis, especially at the discourse level, is done manually by human researchers. This talk explores how the findings and techniques of computer-mediated discourse analysis (CMDA), a paradigm I have been developing and teaching for 18 years, can inform computational approaches to communication in social media. I start by reviewing established automation approaches, which mainly focus on structural linguistic phenomena, and emergent approaches, such as machine learning models that identify semantically- and pragmatically-rich phenomena, through the lens of CMDA, pointing out the strengths and limitations of each. The basic problem is that patterns in the discourse of social media users can be identified by humans that do not appear to lend themselves to reliable automated identification using existing approaches. To begin to address this problem, I draw on examples of recent work on Twitter, Wikipedia, and web-based discussion forums to suggest an approach that synthesizes linguistically-informed manual analysis and existing automated techniques. I consider how such an approach could scale up, while still making use of human analysts, and I identify a number of real-world problems that automated CMDA could help address.