

Complementing Digital Text Data: Design Goals of Multi-modal and Multi-method Approaches

18 August 2022, 20:00-21:20 GMT+8

Organizer & Moderator: Dr. Yuan Hsiao
Assistant Professor, University of Washington, USA

The availability of digital data has led to tremendous increases in research that utilizes text-as-data. Nonetheless, scholars are realizing that textual information can only capture limited dimensions of human behavior, and researchers should complement digital textual data with other forms of data sources. Such sources can be primary data such as surveys or experiments, or they can come from secondary data such as audience engagement data.

The critical question for incorporating multiple data sources is how to best consolidate multiple forms of data. This panel highlights the importance of design goals. The most common design goal of combining data sources is “triangulation”, where different sources of data are used to provide answers to the same question. Another goal is that of “development”, where one data is utilized to develop the data instruments of another source of data. Finally, one should draw attention to the design goal of “sample selection”, where one source of data informs the research scope of another source of data.

This panel brings together scholars that use different design goals. Dr. Daniel Karell from Yale University provides an excellent example of triangulation. Dr. Karell pulls together text and image data, and develops a multi-model framework in the context of security forces. Dr. Karell integrates methods of relational and discursive analyses that provide an easy-to-implement automated approach useful for analyzing the multi-model material underlying meaning-making in many instances of digital discourse.

Dr. Jin Woo Kim from the University of Iowa demonstrates the development. Eying on the relationship between social media and conflict dynamics, Dr. Kim draws from Facebook posts and comments to examine the ranking algorithmic effects of Facebook on the “outrageousness” of each news article. These results then inform the development of an original survey experiment, which tests the downstream effects of ranking systems on affective polarization with US adults.

Panel Session 2

Dr. Yunkang Yang from Texas A&M University pinpoints the nature that the distribution of attention of social media posts is highly skewed, with a few top articles, videos, or comments often getting the most engagement and the rest receiving little. Dr. Yang shows how audience engagement data can be used for the goal of “sample selection”, and researchers can effectively reduce the corpus size for computational textual analysis such as STM by focusing on high engagement content only. Dr. Yang’s approach can be applied to a wide range of online content that involves engagement metrics.