Formally representing uncertainty with scholarly claims for interdisciplinary synthesis: An empirical investigation

Joel Chan, Katrina Fenlon, and Wayne Lutters
College of Information Studies, University of Maryland, College Park

Synthesis is critical for scholarly progress. Effective synthesis requires more than mere "counting" of findings: instead, it involves sophisticated reasoning over the *uncertainty* behind each claim, and how they combine into an overall picture on some problem. For example, the recent shift in understanding of the importance of aerosols as a transmission route for SARS-COV2 relied on synthesis of claims from numerous, partial lines of evidence from diverse disciplines, including epidemiology, physics, virology, and clinical case studies. This kind of synthesis is currently an arduous and slow process, with significant potential for computational support. Here, we focus on the conceptual modeling task of representing uncertainty behind scholarly claims to support computationally-augmented synthesis, such as quantitative aggregation of claims. There are several conceptual difficulties with this task. For example, what constitutes uncertainty, and how it relates to stronger or weaker evidence can be contentious, especially in settings with a high degree of interdisciplinarity. Further, explicit articulation and negotiation of degrees of uncertainty, particularly across epistemological boundaries, might be where true interdisciplinary progress happens; prematurely compressing these nuances into a single context-less number could severely limit the usefulness of that number for synthesis. Yet, ignoring quantitative expressions entirely could significantly limit the scale and complexity of synthesis. To gain more insight into this problem and design opportunities for resolution, we will conduct a participatory observation study of interdisciplinary research groups who are in the process of implementing computational augmentation of their synthesis process, including formal representation of claims and their uncertainty. We will focus particular attention on observing the consequences of key representational design choices (e.g., eschewing quantitative representations of uncertainty entirely, preserving or leaving out provenance of quantitative expressions of uncertainty) for the speed and quality of the synthesis process. We will also document the diversity of artifacts and traces (e.g., notes, snippets from source documents) that may be necessary to appropriately contextualize quantitative expressions of uncertainty from heterogeneous sources. We believe this investigation will yield valuable empirical and design insights to complement the significant formal modeling work that has already explored this question of uncertainty and synthesis.