Accelerating Chemical Assessments: A Case Study in Automatic Evidence Extraction from Text

Catherine Blake¹ and Jodi Flaws²

¹ School of Information Sciences and Department of Computer Science, ² Department of Comparative Biosciences, College of Veterinary Medicine, University of Illinois at Urbana-Champaign.

Abstracts were collected from an earlier study (Korhonen et al. 2012) that characterized evidence and modes of action for 7 chemicals: 4-aminobiphenyl, asbestos, ethylene oxide, pyridine. This pilot study considers two of the chemicals: DEN and WY. The manual processes used to extract mechanistic evidence from studies is one of the most time consuming steps when conducting a chemical assessment. Our goal is to automate evidence extraction in order to reduce the time to conduct a review and/or increase the scope of a review.

Explicit claims involve an agent, the nature of the change, and an object (Blake, 2010). Explicit claims that include a keyword associated with cell proliferation and cell death were identified automatically. Claims were characterized as supporting if they increased the MOA, neutral if no directionality was provided and refuting if they decreased the MOA. Negation was detected.

Explicit claims were identified from abstracts (n=3078) collected in a previous study (Korhonen et al. 2012). Prior mode of action annotations were used to identify keywords using Shannon’s measure of entropy and expert review. The number of supporting claims, where the MOA has increased (e.g. improve, extend), where there was some effect but the claim is neutral (e.g. change, effect), and where the MOA has decreased (e.g. reduce, inhibited) are reported.

Results show that simply reporting a mode of action should not be interpreted as evidence that the MOA has increased. Explicit claims from the Claim Framework provide the granularity necessary to differentiate between supporting, neutral, and refuting claims for a given MOA. Further work is required to differentiate between claims made as background knowledge and the results from current experiments.

Works Cited

Acknowledgments
This material is based upon work supported by the National Science Foundation under Grant No. 1535167.