Literature-based cheminformatics for research in chemical toxicity

Nancy C. Baker[1], Thomas Knudsen[2], Antony J. Williams[2]

[1]Leidos, RTP, NC; [2]National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Agency, RTP, NC

PubMed is the largest freely available source of published literature available online with access to 27 million citations (as of October 2017). Contained within the literature is an abundance of information about the activity of chemicals in biological systems. Literature informatics approaches applied to chemical toxicity research can help researchers make use of the literature in highly effective ways. At the EPA's National Center for Computational Toxicology, in addition to our text-mining capabilities, we have developed a novel approach to article retrieval in our PubMed Abstract Sifter. The Abstract Sifter is a document retrieval tool that integrates the richness of PubMed with the powerful data-manipulation capabilities of Microsoft Excel. Results from a PubMed search are imported directing into an Excel sheet where the end-user can then use a novel "sifter" methodology for quick, agile relevance ranking of articles. The tool also enables article triage capabilities through easy tagging and noting functionality. Triaged citations can be exported to external software such as reference management tools. The Abstract Sifter can also provide a high-level view of a corpus of literature for a defined set of entities such as chemicals. This "landscape" view helps researchers assess the volume of literature in any given subject area to help with project scoping and chemical ranking. A version of the Abstract Sifter is also implemented in the web-based CompTox Chemistry Dashboard developed by the EPA's National Center for Computational Toxicology. Both versions of the tool will be demonstrated and discussed. *This abstract does not necessarily represent U.S. EPA policy.*