

# Text Analytics Suite for Consumer Product Safety Surveillance

**Nominees:** Alan Abrahams, Weiguo (Patrick) Fan, G. Alan Wang, Rich Gruss (*Virginia Tech*)

## Summary

(499 words)

In this age of burgeoning online sales, regulators, such as the Consumer Product Safety Commission (CPSC), National Highway Transportation Safety Commission (NHTSA), and Food and Drug Administration (FDA), face daunting product safety surveillance challenges, that cannot be addressed by traditional physical product testing alone. Online consumer reviews and discussions represent an increasingly viable mechanism for rapid, crowd-sourced quality intelligence.

The research effort nominated here has developed and evaluated a suite of text analytics tools and dictionaries for product safety surveillance from online consumer-generated content. This suite of tools rests firmly in the field of information systems as it provides quality management support through software for high-scale collaborative tagging. The text analytics suite offers machine intelligence for scoring of consumer postings, and facilities for issue identification and management. This research was amongst the first efforts to establish the usefulness of online consumer postings for enterprise and regulatory safety management exercises. Important and notable findings from the development and implementation were that: *traditional sentiment analysis is inadequate for detecting product hazards*, which are frequently stated in non-emotive terms (Abrahams, Jiao, Wang, Fan, 2013; Abrahams, Fan, Wang, Zhang, Jiao; 2015); and that industry-specific smoke term lists are insufficient, and more *fine-grained product-category-specific smoke term lists are necessary for effective hazard discovery*. The research effort has developed and tested custom-tailored procedures and "smoke lists" for hazard identification in a variety of industries, including automobiles (Abrahams, Jiao, Wang, Fan, 2013), consumer electronics (Abrahams, Fan, Wang, Zhang, Jiao; 2015), toys (Winkler, Abrahams, Gruss, Ehsani, 2016), home appliances (Law, Abrahams, Gruss; 2017; Goldberg and Abrahams, 2017), and pain therapies and devices (Adams, Gruss, and Abrahams, 2017).

The research effort has produced and tested a suite of software artifacts: *PamTag* (a highly-scalable web-based collaborative tagging system), *PamTAT* (a text analytics add-in for Excel), and *PamFlag* (a web-browser extension for highlighting, flagging, and logging safety hazards). These software artifacts have been iteratively developed. Features and scoring-dictionaries have been identified, codified, implemented, assessed, and improved, in the research program's published research. The system holds more than 80 million online consumer reviews and discussion. Since 2013 more than one thousand undergraduate and graduate students have conducted data analytics work using this text analytics suite: 30+ industries have been studied, 500,000+ unique online reviews have been tagged, and 1,000+ product safety concerns have been identified. Results have been requested by, and provided at no cost to, dozens of users at large corporations and government regulators, including: *Amazon*, The world's largest retailer; *Intertek*, The world's largest tester of consumer goods; *US CPSC*, The United States Consumer Product Safety Commission; *Health Canada*, the department of the government of Canada with responsibility for national public health; and a *United-States-based appliance manufacturer with more than \$500 million in annual revenue*. Our software tools have also been used by 100+ external university users, including users at Purdue University, University of Maryland, and Johns Hopkins University. The research effort has been principally led and driven by university based faculty, staff and students for R&D and educational purposes.