

# **A Quantitative Approach for Similarity Comparison of the Terminologies in Standard Documents**

**Huixing Meng, Kanittha Setthapitayakul and Yan-Fu Li**

Department of Industrial Engineering

Tsinghua University

Beijing, China

liyanfu@tsinghua.edu.cn

## **Abstract**

The application of standards is effective for ensuring the expected function of a system in a specific scenario. A key step of the standard enactment or implementation is to precisely comprehend the terminologies in the standards via literal comparison. Current methods for such comparison usually rely heavily on labor, which is time consuming and error-prone. With the rapid increase in the number of the standard documents, it is essential to develop an approach to automatize the comparison process. In this work, we proposed a methodology for the computerized comparison of the terms and definitions in standards. Based on the standard structures, the methodology is developed in three steps: the PDF (Portable Document Format) text conversion, the terms and definitions extraction, and comparison. (1) According to the PDF types, either scanned or digitally created, we provided corresponding methods for converting the PDF files. (2) Regarding the terms and definitions extraction, we identified content structures and logical elements, and extracted definition sections and terms. (3) For the comparison step, we evaluated the similarities between the terminologies from the semantic and syntactic aspects. Reliability, availability, maintainability, and safety (RAMS) are crucial attributes to evaluate the performance of a system. In the experimental studies, we compared the RAMS standards issued by IEC, IEEE, ISO, and the Society for Automotive Engineering (SAE). The results show that the proposed methodology is capable of evaluating the similarities of the terms and definitions in standards.

## **Biography / Biographies**

**Huixing Meng** is currently a Postdoctoral researcher at the Reliability & Risk Management Laboratory, Department of Industrial Engineering, Tsinghua University, Beijing, China. He received his Ph.D in Computer Science from Ecole Polytechnique, Palaiseau, France in 2018. His current research areas include the system reliability assessment and risk assessment with the application onto energy systems.

**Kanittha Setthapitayakul** was a Master student of Global Manufacturing in the Department of Industrial Engineering, Tsinghua University, Beijing, China.

**Yan-Fu Li** is currently a Professor at Industrial Engineering Department in Tsinghua University. He was a faculty member of the Laboratory of Industrial Engineering at CentraleSupélec, University of Paris-Saclay, France, from 2011 to 2016. His current research interests include RAMS (reliability, availability, maintainability, safety) assessment and optimization with the applications onto various industrial systems. He has led/participated in several projects supported by EU, France and Chinese governmental funding agencies and various industrial partners. He has co-authored over 90 publications, on international journals, conference proceedings and books. He is an Associate Editor of IEEE Transactions on Reliability and a senior member of the IEEE.