Semantic Similarity in Security Regulations

Sarah Elder
NC State University
seelder@ncsu.edu

Hui Guo
NC State University
hguo5@ncsu.edu

Munindar Singh
NC State University
mpsingh@ncsu.edu

Laurie Williams
NC State University
williams@csc.ncsu.edu

ABSTRACT
Security requirements for a product are often influenced by the federal, state, and local laws; organizational policies; and other regulations the product must comply with. Frequently, a single product must comply with multiple different regulatory documents which may or may not contain related statements. *The goal of this research is to facilitate analysis of security regulations by automatically identifying relations between security regulations using natural language processing and machine learning.*

CCS Concepts
• Social and Professional Topics □ Computing / technology policy • Computing Methodologies □ Artificial Intelligence □ Natural Language Processing •

Keywords
Security policy; Natural Language Processing; Data Mining

1. INTRODUCTION
If a product must comply with multiple regulations, and those regulations conflict or otherwise overlap, developers could build the wrong system if they do not identify these areas of overlap in advance and plan to address them. For example, if one regulation states that the data must be encrypted while another regulation states that the data can be stored in either encrypted or plaintext form, the product must clearly store the data in encrypted form to comply with both regulations. However, this means that both statements must be correctly identified in thousands of lines of text for the correct requirement to be identified. *The goal of this research is to facilitate analysis of security regulations by automatically identifying relations between security regulations using natural language processing and machine learning*. To achieve this goal, we must understand what makes security regulations similar. One of the key aspects of this is the notion of "semantic similarity", or the relations between words [1]. These relations are often reduced to pre-defined, domain-specific, graph-like structures; e.g. ontologies and taxonomies. However, in practice these structures often must be supplemented with information from the target corpus or other sources [2]. We begin our research by examining what characteristics of ontologies and taxonomies are most helpful, and how they can be supported by other tools. Ultimately this understanding will be used to automate the process of identifying related security regulations using natural language processing and machine learning techniques to extract information relevant to a software system.

2. RELATED WORK
The existing literature in this area includes several Systematic Literature Reviews focused on taxonomies or ontologies [2] [3] [4]. These provide useful information in understanding and characterizing the extensive literature on security ontologies and taxonomies. However, none of these examine in detail the additional tools used to supplement ontologies and taxonomies in identifying semantic similarity in practical applications.

3. DISCUSSION
While this research is currently ongoing, our review so far has reinforced our existing assumptions. For example, while users of Souag et. al's ontology tool found it user-friendly, the ontology on its own was not seen as likely to be used during in eliciting and modeling security requirements in a practical setting [5]. Similarly, although taxonomies and ontologies are distinct entities, they are often used for similar purposes and hence should be studied together [6]. This poster describes our ongoing literature review as well as our plans for an initial prototype and its evaluation

4. REFERENCES

[1] G. Sridhara, E. Hill, L. Pollock and K. Vijay-Shanker, "Identifying word relations in software: A comparative study of semantic similarity tools," in *the 16th IEEE International Conference on Program Comprehension, 2008.*, 2008.

[2] D. Sánchez, M. Batet, D. Isern and A. Valls, "Ontology-based semantic similarity: A new feature-based approach," *Expert Systems with Applications,* vol. 39, no. 9, pp. 7718-7728, 2012.

[3] A. Souag, C. Salinesi and I. Comyn-Wattiau, "Ontologies for security requirements: A literature survey and classification," in *International Conference on Advanced Information Systems Engineering*, Berlin, 2012.

[4] C. Blanco, J. Lasheras, E. Fernández-Medina, R. Valencia-García and A. Toval, "Basis for an integrated security ontology according to a systematic review of existing proposals," *Computer Standards & Interfaces,* vol. 33, no. 4, pp. 372-388, 2011.

[5] A. Souag, C. Salinesi, R. Mazo and I. Comyn-Wattiau, "A security ontology for security requirements elicitation," in *International symposium on engineering secure software and systems*, 2015.

[6] P. Savolainen, E. Niemela and R. Savola, "A taxonomy of information security for service-centric systems," in *33rd EUROMICRO Conference on Software Engineering and Advanced Applications, 2007*, 2008.