UNIVERSITY of GUELPH

COLLEGE of ENGINEERING AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

MSc Defence

Friday December 20, 2019 at 1 PM in Reynolds, Room 1101

A Semantic-Based Approach to Reduce the Reading Time of Privacy Policies

Jasmin Kaur

Chair: Dr. Xiaodong Lin
Advisor: Dr. Rozita Dara
Advisor: Dr. Charlie Obimbo
Non-Advisory Committee: Dr. Ritu Chaturvedi

Abstract

Privacy policy is a legal document in which the users are informed about the data practices used by the organizations. Past research indicates that the privacy policies are long and hard to understand. They are also known to have incomplete content. Users are not inclined to read the policy as they have to read long policies to find information about data practices of an organization. The solution that we are proposing in this research is to assist users with finding relevant content to their queries using semantic approach.

This thesis presents the development of domain ontology for privacy policies. Natural Language Processing was used to understand the content of the policies and capture vocabulary for the ontology. This vocabulary was further used to build the ontology so that the ontology highlights relevant sentences related to a privacy concern. We validated and evaluated the ontology using different methods: competency questions, data driven, metric based and user evaluation. Results from the evaluation of ontology show that the amount of text to read is significantly reduced as the users have to only read selected text that ranged from 1% to 30% of a privacy policy. The amount of text depended on the query and its associated keywords. This signifies that the time required to read a policy is significantly reduced as the ontology directs user to the right content for a query. This finding was also confirmed by the results of the user study session. The results from the user study session indicated that the users found ontology helpful in finding relevant selected sentences to read as compared to reading the entire policy.