

ABSTRACT

The escalating volume and heterogeneity of data from diverse social media platforms present significant challenges in digital forensic activities, particularly in the analysis and correlation of evidence. This research aims to design and develop a generalized global ontology model, GENOSIS (GENeralized Ontology for Social Media Investigation Support) to address these challenges. The research methodology involved identifying crucial forensic data types through the validation of Competency Questions (CQs) with digital forensic experts, followed by the development of four local ontologies for platform categories: Online Social Networks (OSN), Microblogging Platforms (MP), Media-content Sharing Sites (MCSSS), and Online Forums/Blogs (OFB). Subsequently, a generalization process integrating semantic and lexical matching techniques, along with directed consolidation, was employed to create GENOSIS. A comprehensive evaluation of GENOSIS was conducted. Qualitative validation, through logical consistency testing using the HermiT reasoner and functional testing with 15 CQs using SPARQL, demonstrated that GENOSIS is consistent and capable of accurately answering all relevant investigative questions. Quantitative validation using AgreementMakerLight (AML) to measure the alignment quality between GENOSIS and the local ontologies yielded good to excellent F-measure scores (ranging from 62.10% to 81.20%), with 100% Recall achieved for three out of the four categories. The research findings indicate that GENOSIS successfully provides a unified, semantically rich, and structured knowledge representation for core cross-platform forensic concepts. The developed global ontology offers a potential solution for standardizing and simplifying the analysis of heterogeneous social media forensic data, thereby supporting more effective and efficient digital investigations.

Keywords: Social Media Forensic, Ontology, Data Generalization, Ontology Engineering