Provenance in Scholarly Publications: Introducing the provCite Ontology

Authors : Maria Joseph Israel, Ahmed Amer

Abstract : Our work aims to broaden the application of provenance technology beyond its traditional domains of scientific workflow management and database systems by offering a general provenance framework to capture richer and extensible metadata in unstructured textual data sources such as literary texts, commentaries, translations, and digital humanities. Specifically, we demonstrate the feasibility of capturing and representing expressive provenance metadata, including more of the context for citing scholarly works (e.g., the authors' explicit or inferred intentions at the time of developing his/her research content for publication), while also supporting subsequent augmentation with similar additional metadata (by third parties, be they human or automated). To better capture the nature and types of possible citations, in our proposed provenance scheme metaScribe, we extend standard provenance conceptual models to form our proposed provCite ontology. This provides a conceptual framework which can accurately capture and describe more of the functional and rhetorical properties of a citation than can be achieved with any current models.