Knowledge Extraction from Mixed-Precision Information

Charles Parker, George Garrity

NamesforLife, LLC. (N4L), Michigan, United States

The knowledge of domain experts is integrated from a variety of information sources, ranging from raw text or data to structured and normalized databases (Mixed Precision Information; MPI). In addition to their subject knowledge these experts also possess in-depth knowledge in querying relevant resources using highly specific Subject Language Terminologies. We introduce a novel solution for extracting knowledge from MPI to provide knowledge workers with verifiable interpretations of observations. Our approach combines semantic and semiotic methods to represent information at multiple levels in concept hierarchies. Our adherence to Knowledge Organization and ontology standards (SKOS-XL, OWL, and RDF) provides clean integration and extension points with existing triple stores, ontologies and thesauri. Our method preserves the precision of the original data, using it at appropriate levels of abstraction to support defeasible inference of relations among entities, enabling exact explanations of reasoning. We demonstrate the effectiveness of this approach over a large corpus of literature from the field of microbiology. This method can be applied to any domain that needs to integrate MPI, as applied to Information Extraction (IE), Multiple Classification of entities, inference and reasoning over ontologies and recommendation systems.