

Abstract

SEMWAP: AN ONTOLOGY-BASED INFORMATION EXTRACTION FRAMEWORK FOR SEMANTIC WEB APPLICATIONS

Charles Alfred Douglas Greenidge

The Semantic Web vision is rapidly becoming a mainstream reality but obstacles remain in the way. In particular the adoption of practical Semantic Web Applications and the production of vast stores of ubiquitous meta-data which is needed to allow robust inferencing engines to attain the goals of machine readability of web documents. We propose the SEMWAP framework which facilitates semi-automatic matching of instance data from opaque web databases using lightweight ontology terms. Our framework uses a mix of Information Retrieval, Information Extraction and Natural Language Processing techniques to produce a matching, providing a viable building block for Semantic Web Applications. We also implement our model in a Java prototype system called SODL, establishing a base from which to extract empirical results.

keywords: Semantic Web; Information Extraction; Ontology Matching; Similarity Matrix; Natural Language Processing.