Abstract

Many languages are in serious danger of being lost and as a result, there has been a significant increase in language documentation projects, and also in attempts to preserve language documentation via the use of digital technologies. We are developing a distributed digital library for endangered languages which will contain various data of endangered languages in the forms of text, image, video, audio and include advanced tools for intelligent cataloging, indexing, searching and browsing information on languages and language analysis. As part of this project, we developed two annotation tools, namely OntoElan and LINGOS. OntoELAN, an ontology-based annotation tool, links linguistic annotations to linguistic ontology terms, and saves and retrieves the multimedia annotations with OWL Web Ontology Language. LINGuistic Ontology managEment System (LINGOES) is a framework to enable linguists to take full advantage of the Semantic Web technologies. Together with OntoGloss, a text annotation tool, and an RDF database with versioning and querying capabilities, it allows a linguist to markup any document with classes in one or more ontologies at the morpheme’s level. Textual documents can be in any language as long as they are accessible via a URI (Universal Resource Identifier). The annotated data can be queried across these languages or can be used to annotate other documents. Saving the annotated data in an RDF repository with inference, querying and change management capabilities makes annotations in LINGOES accessible by machines and useful to the wider Semantic Web community.