▶ The LexMeta Metadata Model for Lexical Resources: Theoretical and Implementation Issues

Penny Labropoulou,
Institute for Language and Speech Processing, Athena Research Center, Athens,Greece, penny@athenarc.gr
David Lindemann,
UPV/EHU University of the Basque Country, Spain, david.lindemann@ehu.eus
Christiane Klaes,
TU Braunschweig/Univ. Library, Germany, c.klaes@tu-braunschweig.de
Katerina Gkirtzou,
Institute for Language and Speech Processing, Athena Research Center, Athens,Greece, katerina.gkirtzou@athenarc.gr

Abstract: The paper presents LexMeta, a metadata model catering for descriptions of human-readable and computational lexical resources included in library catalogues and repositories of language resources. We present the main concepts of the model, its implementation, and discuss current findings and future plans.

Purpose: The paper presents LexMeta, a metadata model catering for descriptions of human-readable and computational lexical resources included in library catalogues and repositories of language resources. It, therefore, takes into account requirements for bibliographical metadata (for citation purposes) as well as for features of contents and form, addressing user needs of lexicographers and, in a broader context, of scholars of social sciences and humanities (SSH) that wish to deploy these resources in their research workflows (for findability and usability purposes).

Design/methodology/approach: The model builds upon broadly used existing models; these are mainly the FRBR (Functional Requirements for Bibliographic Resources) model from librarians' perspective, the META-SHARE ontology for language resources and technologies, and the LexVoc Vocabulary of Lexicographic Terms, a structured controlled list of terms related to lexicographical and metalexicographical concepts. New classes and properties are introduced in the LexMeta namespace when these are not covered by existing models. The model is structured around three main classes: *Lexicographic Work*, equivalent to the abstract notion of a lexicographical creation, *Lexical/Conceptual resource*, representing the realization of a single work (e.g., a certain version or edition of a lexicographic work) and *Distribution*, representing the physical form in which a work is issued (e.g., as a printed book or a digital file). For each class, the appropriate set of properties is defined accompanied with cardinality features, and, for classification properties, the relevant controlled vocabularies. Development of the model is underway in two forms: as an ontology of Wikibase entities,

and as an OWL ontology. The former implementation enables its adoption in the LexBib Wikibase Knowledge Graph of Lexicography and Dictionary Research, a research infrastructure that follows the Wikibase data model, while the latter supports works in the Linked Data community. Classification vocabularies are represented in the form of SKOS.

Findings: The model looks promising for the envisaged purposes.

Research limitations/implications: LexMeta is used in the population of the LexBib catalogue, which helps identify potential gaps and errors. We are also soliciting feedback from the target communities.

Practical implications: The model can be extended and enriched with recommendations and user feedback. We foresee issues in keeping the two representation forms synchronized, and we are looking into a viable solution.

Originality/Value: To the best of our knowledge, this is the only attempt at the dual implementation of a metadata model. The metadata model itself is of particular interest to SSH researchers and lexicographers, and is already applied in a real use case.

Keywords: *lexicography, metadata model, lexical resources, linked data, wikibase.*

Research type: Research paper