

Mireille Ducassé/Archil Elizbarashvili

## FINDING LEMMAS IN AGGLUTINATIVE AND INFLECTIONAL LANGUAGE DICTIONARIES WITH LOGICAL INFORMATION SYSTEMS

### The case of Georgian verbs

**Abstract** Looking up for an unknown word is the most frequent use of a dictionary. For languages both agglutinative and inflectional, such as Georgian, this can be quite challenging because an inflected form can be very far from the lemmas used by the target dictionary. In addition, there is no consensus among Georgian lexicographers on which lemmas represent a verb in dictionaries. It further complicates dictionaries access. Kartu-Verbs is a base of inflected forms of Georgian verbs accessible by a logical information system. It currently contains more than 5 million inflected forms related to more than 16 000 verbs for 11 tenses; each form can have 11 properties; there are more than 80 million links in the base. This demonstration shows how, from any inflected form, we can find the relevant lemma to access any dictionary. Kartu-Verbs can thus be used as a front-end to any Georgian dictionary.

**Keywords** E-dictionary; lemma; Georgian language; under-resourced language; Inflected forms; Logical information systems; semantic web

### Contact information

**Mireille Ducassé**

IRISA-INSA Rennes

mireille.ducasse@irisa.fr

**Archil Elizbarashvili**

Ivane Javarishvili Tbilisi State university

archil.elizbarashvili@tsu.ge