LREC 2014 Workshop

CCURL 2014

Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era

Reykjavík (Iceland), 26th May 2014

Poster Presentation

Title

Computerization of African Languages-French Dictionaries

Authors

Chantal Enguehard, Mathieu Mangeot

Abstract

This paper relates work done during the DiLAF project. It consists in converting 5 bilingual African language-French dictionaries originally in Word format into XML following the LMF model. The languages processed are Bambara, Hausa, Kanuri, Tamajaq and Songhai-zarma, still considered as under-resourced languages concerning Natural Language Processing tools. Once converted, the dictionaries are available online on the Jibiki platform for lookup and modification. The DiLAF project is first presented. A description of each dictionary follows. Then, the conversion methodology from .doc format to XML files is presented. A specific point on the usage of Unicode follows. Then, each step of the conversion into XML and LMF is detailed. The last part presents the Jibiki lexical resources management platform used for the project.