

Merging linguistic databases with the Web Ontology Language (OWL)

Sebastian Nordhoff

November 18, 2013

Abstract

1 Introduction

problem many datasets heterogeneous structure different domains phonology, syntax, lexical semantics (geographical) current workflow project collect data from grammars use spreadsheet or local database run analyses publish analyses problems raw data not available raw data in heterogeneous formats different concepts BlakeCysouw different languoids

this paper publish data with RDF merge data cross-query with SPARQL

2 The current state of the art of linguistic databases

3 Issues with the current setup

4 Interoperability

syntactic FileMaker SQL RDF semantic GOLD language, feature, value rdf triples

5 desiderata

cross-query TDS

6 Publish data

publish data lingtyp.owl doculects linguistic descriptables features values annotations

7 Merging data

merge data merge features merge languoids

8 Case studies

case study WALS PHOIBLE ASJP

9 Cross-querying data

cross-query databases which languages are SOV languages have fricatives have
a ASJP distance of 10 or less to Hebrew are spoken in altitudes above 1000m

10 Sample queries

sample queries

11 Outlook

outlook more datasets current projected desired more domains altitude visual-
ization user experience