Using NVH as a Backbone Format in the Lexonomy Dictionary Editor

Miloš Jakubíček, Vojtěch Kovář, Michal Měchura and Adam Rambousek

Lexical Computing
jak@fi.muni.cz
milos.jakubicek@sketchengine.eu

NLP Centre, Masaryk University

RASLAN 2022, 9. 12. 2022

Outline

Background: the world of DWS

- IDM DPS
- iLex
- TLex (TshwaneLex)
- DEB platform
- Lexonomy
- many per-dictionary projects
- Oxygen & Vim

Lexonomy

- originally developed by Michal Měchura, later adopted by Lexical Computing
- small, web-based, easy to customize

Dictionary formats

data models

- LMF (ISO standard)
- TEI Lex-0
- LEXIDMA (upcoming OASIS standard)

serializations

- XML
- JSON
- plaintext
- RDF (and more complex)
- NVH (name-value hierarchy)

On standards

LEXIDMA: motivation

- narrow focus
- free availability
- modern digital dictionary
- not retrodigitization
- not paper production
- real-world scope of lexicography: focus on what is *normal*, not on what is *possible*

NVH: motivation

- easy to read by humans
- not difficult to write by humans
- easy to process by computers
- solving XML bottlenecks (headedness, matryoshkization)
- Měchura (2022): *Better than XML: Towards a Lexicographic Markup Language*

NVH: implementation

- https://www.namevaluehierarchy.org/
- https://github.com/michmech/nvh
- toolchain in Python and JS
- parser, serializer, merger, splitter, schema generation and validation

NVH: examples

https://github.com/michmech/nvh/blob/master/python/python.md

Outline
The world of DWS
Lexonomy and dictionary formats
Conclusions
○
○
○○○○○○○
●

Conclusions

- lexicographic world obsessed with XML(-based) technologies though cannot used them
- LEXIDMA and NVH coming
- also as a backbone for Lexonomy
- also as a backbone for post-editing lexicography