

Validating Danish Wikidata lexemes

Finn Årup Nielsen,¹ Katherine Thornton² and Jose Emilio Labra Gayo³

¹Cognitive Systems, DTU Compute, Technical University of Denmark

²Yale University Library, New Haven, CT, USA

³University of Oviedo, Spain

10 September 2019

Validating Danish Wikidata lexemes

```

start = @<danish-noun>

<danish-noun> {

  dct:language [ wd:Q9035 ]
    // rdf:label "language"
    // rdf:comment "language of lexeme must be Danish" ;

  wikibase:lexicalCategory [ wd:Q1084 ]
    // rdf:label "lexical category"
    // rdf:comment "Lexical category for a noun must be noun" ;

  wdt:P5187 [ @da ]
    // rdfs:label "Danish noun word stem"
    // rdfs:comment "Danish noun must have a Danish word stem" ;

  wdt:P5185 [ wd:Q1305037 wd:Q1775461 ]
    // rdf:label "grammatical gender"
    // rdf:comment "grammatical gender of noun must be either masculine or feminine" ;

  p:P6140 EXTRA a {
    a [ wdno:P6140 ] | ps:P6140 /^[0-9]{8}$/
    // rdf:label "DanNet identifier"
    // rdf:comment "DanNet identifier should either be numeric or alphanumeric" ;
  }
}

```

Wikidata + lexicographic data
+ validation with ShEx

Definition of ShEx schemas
for a range of Danish word
classes: nouns, verbs, pro-
nouns, numerals.

For instance, plural definite
non-genitive forms for Dan-
ish nouns should end with
“er?ne”.

For discovering missing or
wrong information and discuss
data models.