

## ILMĀRS POIKĀNS

LU Matemātikas un informātikas institūts  
ilmars.poikans@lumii.lv

### Latviešu valodas vēsturiskas vārdnīcas izveide TshwaneLex\* sistēmā

Ja vārdnīca tiek veidota, izmantojot teksta redaktoros vai maketēšanas programmas, tad rodas pieraksta un formatējuma konsekvences problēmas, it sevišķi, ja vārdnīcas veidošanās piedalās vairāki cilvēki – sīkas atšķirības formatējumā, neuzmanības kļūdas struktūrā. Taču šādas pieejas lielākais trūkums ir šķirkļa struktūras informācijas zaudēšana. Lai arī ir iespējams apstrādāt formatētu tekstu un automātiski iegūt strukturētu informāciju par šķirkli, tomēr tas ir ļoti darbietilpīgs un kļūdu pilns process.

Ja tiek saglabāta šķirkļu strukturālā informācija, tad veidotāji var automātiski datus pārbaudīt, pārveidot un iegūt jebkuru nepieciešamo izejas formatējumu. Pastāv variants jau sākotnēji marķēt katra šķirkļa struktūru tekstā ar speciāliem simboliem vai tagiem, taču tas ir lietotājiem neērts un apgrūtinošs darbs, kurā var ļoti viegli ielaist kļūdas.

Taču vispiemērotākais līdzeklis vārdnīcas izveidei ir specializēta programma, kas saglabā vārdnīcas šķirkļu strukturālo informāciju, validē dažādu lauku vērtības, samazinot iespēju kļūdīties. Viens no variantiem ir TshwaneLex sistēma. Šīs sistēmas priekšrocības:

Plaši konfigurējama, lai pielāgotu konkrētās vārdnīcas izstrādei un formātam;

Draudzīga lietotājam – nav nepieciešama ilga apmācība;

Daudzlietotāju režīms – nav problēmu ar vairāku cilvēku darba apvienošanu;

Tūlītējs šķirkļa attēlojuma priekšskats;

Datu eksports uz XML un HTML.

Referātā tiks apskatīta iegūtā pieredze, sastaptās problēmas un iespējamie risinājumi, izmantojot TshwaneLex sistēmu, kura tiek izmantota Latviešu valodas vēsturiskās vārdnīcas izveidē.

\* TshwaneLex – <http://tshwanedje.com/tshwanelex/>