

Pētniecības projekta Nr. 13  
«Meža nozares komunikācijas tehnoloģijas»  
(projekta ID: 1.2.1.1/16/A/009)  
Starpposma 01.01.-31.03.2018 rezultāta atskaite

SIA «TILDE»

31.03.2018

# 4. starpposma rezultāts

## **Starprezultāta īss apraksts:**

1. Turpinājās darbs pie meža nozares tekstu korpusu apzināšanas un vākšanas
2. Pētītas dažādas neironu tīklu mašintulkošanas arhitektūras un to piemērotība projekta specifikai.

## 4. starpposma rezultāts

### **Mašintulkošanas treniņdatu vākšanas aktivitāte**

Identificēts potenciāls, liels satura avots *EU Bookshop*, kas varētu saturēt arī ar mežu jomu saistītu saturu. Iesākta satura kataloga un metadatu savākšana un analīze, lai identificētu konkrētos ar mežu nozarei saistītos dokumentus tajā. No *EU Bookshop* tiek paredzēts iegūt ap 200 tūkstošus paralēlu teikumu mežu nozares mašintulkošanas treniņdatu. Aktivitāte beigsies nākamā kvartālā.

# 4. starpposma rezultāts

## **Mašintulkošanas sistēmu izveides aktivitāte**

*Pētītas dažādas neironu tīklu mašintulkošanas arhitektūras un to piemērotība projekta specifikai, t.i., tulkošanai latviešu valodā, kas ir morfoloģiski bagāta, un jomā, kurā ir pieejams neliels treniņdatu apjoms*

*Balstoties uz pētījuma rezultātiem, sagatavots un iesniegts publicēšanai raksts LREC 2018 konferencē:*

Rikters, M., Pinnis, M., & Krišlauks, R. (2018). Training and Adapting Multilingual NMT for Less-resourced and Morphologically Rich Languages. In N. C. (Conference Chair), K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, ... S. Piperidis (Eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Paris, France: European Language Resources Association (ELRA).