

Pētniecības projekta Nr. 13
«Meža nozares komunikācijas tehnoloģijas»
(projekta ID: 1.2.1.1/16/A/009)
Starpposma 01.04.-30.06.2018 rezultāta atskaite

SIA «TILDE»

30.06.2018

5. starpposma rezultāts

Starprezultāta īss apraksts:

1. Pabeigts darbs pie meža nozares tekstu korpusu apzināšanas un vākšanas
2. Izveidota meža nozarei pielāgotā neironu tīklu mašintulkošanas sistēma.

5. starpposma rezultāts

Mašīntulkošanas treniņdatu vākšanas aktivitāte

Pabeigts darbs pie meža nozares tekstu korpusu apzināšanas un vākšanas. Kopā mašīntulkošanas sistēmu trenēšanai sagatavoti 233 tūkstoši teikumu nozares tekstu un 12 miljoni teikumu vispārīgu tekstu. Sagatavoti arī sistēmu pielāgošanas un novērtēšanas dati, attiecīgi 2000 un 1000 paralēlu teikumu.

5. 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. Izveidota bāzes mašintulkošanas sistēma un tā pielāgota nozarei, izmantojot projektā savāktos meža nozares tekstus. Veikta sistēmas kvalitātes automātiskā novērtēšana.

Balstoties uz pētījuma rezultātiem, sagatavoti publicēti 2 raksti:

Pinnis, M., Skadiņš, R., Šics, V., & Miks, T. (2018). Integration of neural machine translation systems for formatting-rich document translation. In M. Silberztein, F. Atigui, E. Kornysheva, E. Métais, & F. Meziane (Eds.), *Natural Language Processing and Information Systems. NLDB 2018. Lecture Notes in Computer Science* (Vol. 10859, pp. 494–497). Paris, France: Springer, Cham. http://doi.org/10.1007/978-3-319-91947-8_51

Pinnis, M., Vasiljevs, A., Kalniņš, R., Rozis, R., Skadiņš, R., & Šics, V. (2018). Tilde MT Platform for Developing Client Specific MT Solutions. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, ... T. Tokunaga (Eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (pp. 1344–1348). Miyazaki, Japan: European Language Resources Association (ELRA). Retrieved from <http://www.lrec-conf.org/proceedings/lrec2018/summaries/8882.html>