

Pētniecības projekta Nr. 13  
«Meža nozares komunikācijas tehnoloģijas»  
(projekta ID: 1.2.1.1/16/A/009)  
Starpposma 01.07.-30.09.2017 rezultāta atskaite

SIA «TILDE»

30.09.2017

## 2. Starpposma rezultāts

### Starprezultāta īss apraksts:

- **Apzināti** Meža nozares pārstāvji un darbojošās juridiskās personas (tā skaitā mācību iestādes, uzņēmumi un citi)
- **Identificēti** datu avoti Meža nozares valodas resursiem
- **Izgūti** dati no Latvijas tiešsaistes avotiem (monolingvālie un bilingvālie dati) izmantojot datu izgūšanas rīkus un darbplūsmas
- **Apstrādāti** dati izmantojot datu apstrādes rīkus (tajā skaitā domēnu datu atdalīšana/filtrēšana no vispārēja rakstura tekstiem)
- Iesākta monolingvāla un bilingvāla korpusa **sagatavošana** Meža nozares domēnā, ko tālāk izmantos mašīntulkošanas sistēmu trenēšanai

## 2. Starpposma rezultāts

Papildus informācija par sasniegto rezultātu t.sk.:

- Identificētie datu avoti Meža nozares valodas resursiem:
  - Valsts mēroga ar mežu un dabu saistīto institūciju mājas lapu
  - Meža jomas zinātnes un izglītības vietnes tīmeklī
  - Meža apstrādes un mežsaimniecības industrijas uzņēmumu mājas lapas
  - Tūrisma un dabas rezervātu mājas lapas
  - Meža nozares enciklopēdijas un terminoloģijas vārdnīcas
- Kopā izgūti un apstrādāti **39 000 monolingvālie segmenti** un **2000 bilingvālie segmenti** Meža nozares tekstiem