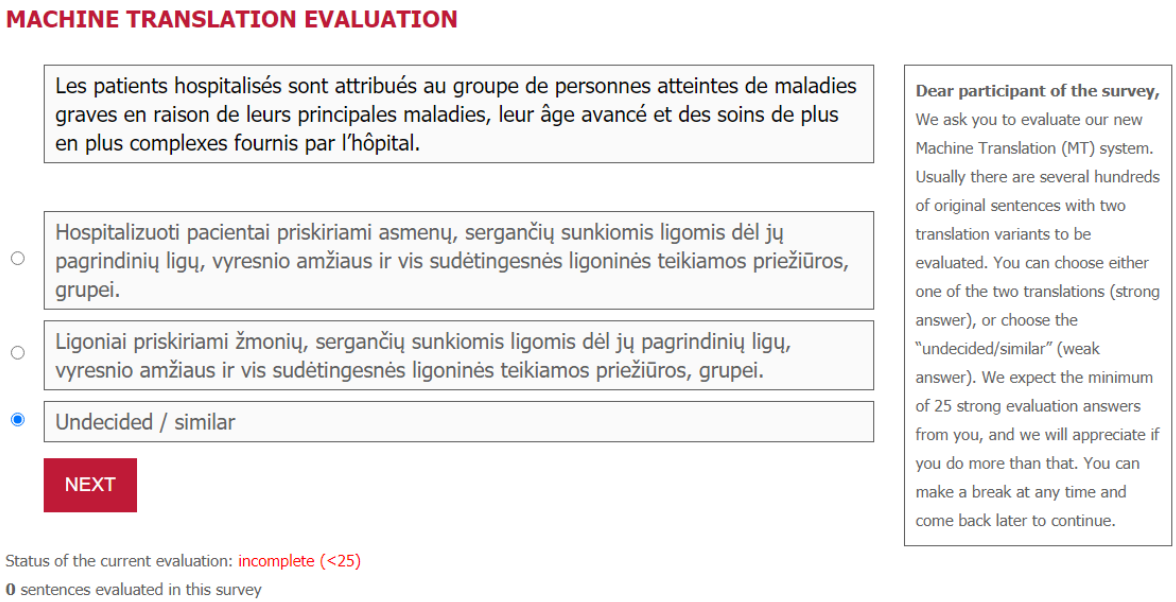
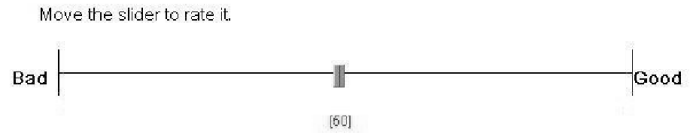
Prakses iespēja Tildē: Mašīntulkošanas rezultātu kvalitātes cilvēkvērtēšanas tiešo vērtējumu rīkkopas izstrāde

Mašīntulkošanas sistēmu izstrādē iespējams svarīgākais posms ir rezultātu vērtēšana. Praksē tiek izmantotas dažādas automātiskās metrikas, taču kā svarīgākais joprojām tiek uzskatīts cilvēku, it īpaši profesionālu tulkotāju vērtējums. Jau ilgāku laiku Tilde MT iekšējos procesos lieto rīku, kas ļauj salīdzināt divu sistēmu tulkojumu savstarpējo kvalitāti – sistēma A labāka par sistēmu B, B labāka par A, vai A un B vienlīdz labas/sliktas.



Tomēr, šāda rezultātu vērtēšana ir neērta, līdzko jāsalīdzina vairāk nekā divas sistēmas. Turklāt kategoriskie salīdzinājumi "A labāka par B", neļauj atbildēt uz jautājumu, cik daudz viena sistēma ir labāka par otru.

Prakses laikā tiks izstrādāta rīkkopa mašīntulkošanas rezultātu tiešajai vērtēšanai, kurā vērtētājiem tiks lūgts vērtēt teikuma kvalitāti nepārtrauktā Likerta skalā (https://en.wikipedia.org/wiki/Likert_scale):



Papildus vērtēšanas funkcionalitātei, projektā paredzēts izstrādāt arī rezultātu statistiskās analīzes funkcionalitāti, kas būtu piemērota tiešās vērtēšanas metodoloģijai.

RESULTS OF MACHINE TRANSLATION EVALUATION

Logged in as [redacted]

Source - target language:	EN - LT
Description:	[redacted]
Date created:	9/9/2020 4:01:13 PM
Now:	11/9/2020 5:20:23 PM
Is quantitative:	YES
Escape special chars:	YES

System 1

Name:	[redacted]NMT
Url:	[redacted]
Bleu score:	23.06

System 2

Name:	Google translation
Url:	https://translate.google.com/
Bleu score:	21.90

Calculation results

Conclusion:	System 1 is better according to the normal approximation interval
Credibility:	Weakly sufficient
Fleiss' kappa interpretation:	0.331 (Fair agreement)
Free kappa interpretation:	0.338 (Fair agreement)
Export data:	CSV

ALL EVALUATIONS

System 1:	159
System 2:	119
Tie:	146
Total evaluations:	424

Name	p ± ci	lower	upper
System 1	37.50% ± 4.61%	32.89%	42.11%
System 2	28.07% ± 4.28%	23.79%	32.34%
Tie	34.43% ± 4.52%	29.91%	38.96%

All evaluations ignoring ties

Name	p ± ci	lower	upper
System 1	57.19% ± 5.82%	51.38%	63.01%
System 2	42.81% ± 5.82%	36.99%	48.62%

Rīkkopa tiks realizēta kā tīmekļa pakalpojums (web service), kur augstāk minētā funkcionalitāte pamatā tiks īstenota serverpusē. Tiks izstrādāta vienkārša funkcionāla lietotāja saskarne tīmekļa pārlūkprogrammā. Serverpuses kods būs jāizstrādā C# vai Python programmēšanas valodā (atkarībā no kandidāta vēlmēm varam izskatīt arī F#). Kandidātam ir vēlamas zināšanas kādā no šīm valodām.

Kandidātiem būs jāpiedalās darba intervijā prasmju noteikšanai. Prakse notiks sabiedrības Tilde Mašīntulkošanas Grupā.

Praksi vadīs Rihards Krišlauks (rihards.krislauks@tilde.lv), konsultēs Toms Bergmanis un Mārcis Pinnis.