



Motivācija

Dabiskās valodas izpratnes (*NLU*) uzdevumi tiek izmantoti jau kādu laiku – lietotāja ievades saprašanas algoritmi (*WolframAlpha*), tulkošana, teksta analīze, balss-asistenti, u.t.t. Lai mašīn-mācīšanās algoritmi šos uzdevumus varētu veikt efektīvāk ir nepieciešami attēlojumi, jeb teksta sintaktiskie un semantiskie modeļi.

Šī darba motivācija ir uzlabot Latviešu valodas resursu <https://github.com/LUMII-AILab/FullStack>, papildinot to ar abstraktās nozīmes reprezentācijas (*AMR*) attēlojumu, rezultātā (cerams) dodot iespēju lielākam un labākam *NLU* risinājumu klāstam Latviešu valodā.

Darba autors: Rihards Sprukulis

Stud. apl. nr.: rs18092

Darba vadītājs: Dr. Dat. Normunds Gruzītis

Problēma

Nepieciešamas lielas, kvalitatīvas tekstu kopas (korpusi) ar attēlojumiem, lai risinātu *NLU* uzdevumus. Manuāli veidot attēlojumu ir dārgi, sarežģīti un ilgi, bet automātiski – nav pietiekoši kvalitatīvi.

Mērķi

Kursa darba mērķis ir literatūras avotu analīze: apskatīt pieejas, kā var iegūt *AMR* attēlojumu un nodrošināt tā kvalitāti.

ABSTRAKTĀS NOZĪMES REPREZENTĀCIJAS ATVASINĀŠANA NO ESOŠAJIEM SINTAKTISKAJIEM UN SEMANTISKAJIEM ATTĒLOJUMIEM LATVIEŠU VALODĀ

Uzdevumi

Iegūt priekšstatu par pieejamiem Latviešu valodas attēlojumu resursiem (*LUMII-AILAB*).

Analizēt jau eksistējošus vai līdzīgus risinājumus.

Apskatīt kvalitātes noteikšanas pieejas ar mērķi pārliecināties par iespējamā rezultāta praktisku ieguvumu.

Secinājumi

Izmantojot automātiskus parsētājus var iegūt nepieciešamo rezultātu, bet ir vajadzīgs parsētājs Latviešu valodā vai jāreķinās ar automātiskās tulkošanas iesaisti, kas var samazināt attēlojuma kvalitāti.

Attēlojuma kvalitātes noteikšanai var izmantot gan salīdzināšanu ar kvalitatīvu attēlojumu, gan praktiska rezultāta pārbaudi, lai pārliecinātos, ka rezultāts ir labāks, par automātisko pieeju ar tulkošanas iesaisti.

Rezultāti

Analizēts Latviešu valodas resurss un tajā izmantotie attēlojumi.

Analizēti sekojoši eksistējošie risinājumi:

izvešana no citas valodas,

parsētāju izmantošana,

attēlojumu izvešana no *part-of-speech* attēlojuma.

Apskatīti divi veidi, kā var noteikt attēlojuma kvalitāti:

salīdzināšana ar cilvēka izveidoto attēlojumu (*SemBleu*, *Smatch*)

Dabiskās valodas izpratnes uzdevumu rezultātu salīdzināšana (*Bleu*)

Turpmākie uzdevumi

Noteikt, vai ir labāka pieeja, kā var izvest attēlojumu (piem., no *FrameNet*).

Noteikt optimālo pieeju atvasināšanai.

Veikt atvasināšanas implementāciju, atvasināšanu un rezultāta kvalitātes analīzi.