

Ajaseoste automaatne tuvastamine tekstis

Juhendaja: Siim Orasmaa (*siim . orasmaa @ ut . ee*)

Ajasemantika märgendus toob loomuliku keele tekstis esile sündmusviited, ajaväljendid ja nende vahelised ajaseosed. Näide:

Esimene teade [põlengust] [saabus] häirekeskusesse [esmaspäeva öösel kell 2.31]. Kui esimene tuletõrjeauto kohale [jõudis], oli hoidla katus sisse [langenud] ning suurem leek [sumbunud].

Ajaseosed:

```
[põlengust] INCLUDES [saabus]
[saabus] SIMULTANEOUS [esmaspäeva öösel kell 2.31]
[jõudis] AFTER [saabus]
[jõudis] AFTER [langenud]
[jõudis] AFTER [sumbunud]
```

Lõputöö eesmärgiks on luua prototüüp süsteemist, mis tuvastab automaatselt eestikeelsestest (ajalehe)tekstidest sündmusviited ja neid hõlmavad ajaseosed. Selleks tuleb kasutada juhendatud masinõpet ning treenida Eesti TimeML korpusel välja vastavad mudelid. Prototüüp tuleks implementeerida Pythonis, toetudes [EstNLTK](#) raamistikule. Andmehulga suuruse tõttu tuleb tõenäoliselt rakendada traditsioonilisi masinõppe algoritme (nt teegi *scikit-learn* <http://scikit-learn.org> vahendeid), aga katsetada saab ka [BERT-i eeltreenitud mudelite](#) kaasamist protsessi.

Viiteid:

- Eesti TimeML korpus: <https://github.com/soras/EstTimeMLCorpus>
- Pustejovsky, J., & Verhagen, M. (2009). SemEval-2010 task 13: evaluating events, time expressions, and temporal relations (TempEval-2). In Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (pp. 112-116). Association for Computational Linguistics.
- Verhagen, M., Sauri, R., Caselli, T., & Pustejovsky, J. (2010). SemEval-2010 task 13: TempEval-2. In Proceedings of the 5th international workshop on semantic evaluation (pp. 57-62). Association for Computational Linguistics.