

Automaatne sisukokkuvõtete tegemine ajaseoste märgenduste põhjal

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

Automaatse sisukokkuvõtete tegemise eesmärgiks on luua tekstist lühendatud versioon, mis sisaldab vaid kasutajale tarvilikku informatsiooni. Lühendatud ja informatiivsed ülevaated on väga olulised suure hulga tekstiliste andmete organiseerimisel ja hõlpsamalt loetavamaks muutmisel, nt leiavad need rakendust interneti otsingumootorites, mobiiltelefonide või pihuarvutite ekraanidel (Müürisep, 2006). Käesoleva töö eesmärgiks on uurida, kuidas saab rakendada ajaseoste märgendust automaatsel sisukokkuvõtete tegemisel.

Planeeritavad alametapid:

- 1) Tekitada ajasemantiliselt märgendatud korpuse (Eesti TimeML korpus) põhjal sisukokkuvõtete tegijale treening- ja testkorpus. Selleks tuleb tekstid läbi vaadata ning valida käsitsi välja sisukokkuvõteteks sobivad laused;
- 2) Joondada korpus selle algallikaga (XML failides) ning lisada korpusele algallika põhjal puuduolev tekstivormingu märgendus (lõigupiirid, kaldkirjas / rasvane tekst);
- 3) Eksperimenteerida korpusel erinevate sisukokkuvõtete tegemise meetoditega. Re-implementeerida varasem lähenemine, mis toetub informatsioonile lausete asukoha, vormingu ja sõnavara kohta (Müürisep 2006), ning luua oma lähenemine, mis kasutab ära ajasemantilisest märgendusest saadavat informatsiooni. Inspiratsiooniks ajasemantilise märgenduse kasutamise kohta vt nt Liu jt (2009);

Vt ka varasemaid töid eesti keeles automaatse sisukokkuvõtete tegemise vallas: Müürisep (2006) ja Sellik (2008).

Viited

Eesti TimeML korpus:

<http://keeleressursid.ee/et/keeleressursid-cl-ut/korpused/83-sample-data-articles/clutee-lehed/276-ajasemantiliste-margendustega-tekstikorpus>

Maofu Liu, Wenjie Li, Huijun Hu. Extractive Summarization Based on Event Term Temporal Relation Graph and Critical Chain. Information Retrieval Technology. Springer Berlin Heidelberg, 2009. 87-99.

Kaili Müürisep. Eestikeelsete tekstide sisukokkuvõtjast EstSum. Keel ja arvuti. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 6. (Toim. M. Koit, R. Pajusalu, H. Õim) Tartu 2006. lk 115-125.

<http://uuslepo.it.da.ut.ee/~kaili/papers/myyrisepsisukokku.pdf>

Keili Sellik. Automaatse sisukokkuvõtja töö hindamine. Bakalaureusetöö. 2008.

http://lepo.it.da.ut.ee/~kaili/juhendamised/Baka_Sellik.pdf